EMPIRICAL BAYES ESTIMATION OF THE RESPONSE FUNCTION AND MULTIVARIATE REGRESSION MODEL

By

LI-CHU LEE

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1989

ACKNOWLEDGEMENTS

The author would like to express her sincere gratitude to her advisor, Dr. Malay Ghosh, for his ideas, advice and support throughout this study. The author is also grateful to Dr. Michael A. DeLorenzo, Dr. Ramon C. Littell, Dr. Kenneth M. Portier, and Dr. Pejaver V. Rao for their participation as members of her graduate committee.

Some valuable data and suggestions were provided by Dr. Ramon C. Littell. His help is gratefully acknowledged. The author wishes also to thank Dr. Kenneth M. Portier and Mr. Doug Reese for their programing suggestions.

Finally, the author would like to thank her husband Li-Hwa Lin and her parents for their support and encouragement.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

LIST OF TABLES v

ABSTRACT vi

CHAPTERS

1 INTRODUCTION 1

1.1 Background 1

1.2 Outline of Work 8

2 EMPIRICAL AND HIERARCHICAL BAYES ESTIMATION OF THE RESPONSE FUNCTION 13

2.1 Introduction 13

2.2 The Empirical Bayes Approach 15

2.3 The Hierarchical Bayes Approach 33

2.4 Numerical Examples 39

3 EMPIRICAL BAYES ESTIMATION OF MULTIVARIATE REGRESSION MODEL I 52

3.1 Introduction 52

3.2 The Empirical Bayes Approcah 54

3.3 Empirical Bayes Subset Estimators 72

3.4 Numerical Examples 81

4 EMPIRICAL BAYES ESTIMATION OF MULTIVARIATE REGRESSION MODEL II 87

4.1 Introduction 87

4.2 The Empirical Bayes Approach 89

4.3 Empirical Bayes Subset Estimators 106
4.4 Numerical Examples 113
5 CONCLUSIONS 119
5.1 Summary and Conclusions 119
5.2 Future Studies 121
BIBLIOGRAPHY 122
BIOGRAPHICAL SKETCH 124

LIST OF TABLES

2.1 The estimated frequentist risk results for 3^2 factorial experiment. 41

2.2 The estimated frequentist risk results for 2^3 factorial experiment replicated two times. 42

2.3 The estimated Bayes risk results for 3^2 factorial experiment. . . 44

2.4 The estimated Bayes risk results for 2^3 factorial experiment replicated two times. 45

2.5 Six-week heights of chrysanthemum plants in pots treated with two growth retardant chemicals at six rates of application. 46

2.6 Sum of squared errors and the Euclidean distances for different estimators. 48

2.7 Predicted values for different estimators. 50

3.1 The estimated frequentist risk results for 2^3 factorial experiment. 83

3.2 The estimated Bayes risk results for 2^3 factorial experiment. . . 85

4.1 The estimated frequentist risk results for 2^3 factorial experiment. 115

4.2 The estimated Bayes risk results for 2^3 factorial experiment. . . 117

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

EMPIRICAL BAYES ESTIMATION OF THE RESPONSE FUNCTION AND MULTIVARIATE REGRESSION MODEL

By

LI-CHU LEE

May 1989

Chairman: Dr. Malay Ghosh
Major Department: Statistics

In the analysis of data from designed experiments, often the problem is to estimate the response function. In doing so, there is often a choice between two nested regression models, for example, one involving the main effects and the interactions, and the other involving the main effects only. Another example involves one model having intercept, linear and quadratic terms, while the other model has only intercept and linear terms. We consider empirical Bayes and hierarchical Bayes estimators of the response function in univariate regression models and empirical Bayes estimators of the response function in multivariate regression models.

The proposed estimators are compromises between the least squares estimators for the full and reduced models, as well as between the least squares estimator for the full model and a weighted average of the least squares estimators for the reduced model. The shrinkage coefficients of these estimators are data dependent. Several theorems are given which describe the frequentist as well as the Bayesian properties of these proposed estimators. Some modified empirical Bayes estimators are developed and compared with the prelimary test estimators. Numerical examples are given which illustrate the estimation methods.

CHAPTER 1
INTRODUCTION

1.1 Background

In his seminal 1956 Berkeley Symposisum paper, Stein presented the surprising phenomenon that the sample mean was an inadmissible estimator of the normal population mean under squared error loss in dimensions higher than two. Later, James and Stein (1961) produced an explicit estimator dominating the sample mean. Since the sample mean is a minimax estimator of the normal population mean under squared error loss, any estimator dominating the sample mean is also minimax. Since then many articles have been devoted to finding a class of minimax estimators which dominate the sample mean. Mention should be made of the articles of Baranchick (1970), Strawderman (1971), Bock (1975), Efron and Morris (1976), Berger (1976) and Berger (1982) among others.

In the seventies, Efron and Morris, in a series of articles, gave an interesing empirical Bayes interpretation of the James-Stein estimator. An empirical Bayes estimator is an estimator that the unknown parameters of the corresponing Bayes estimator are estimated from the data. This opened up the possibility of applying the James-Stein estimator to many data-analysis problems, whenever there was a need to 'borrow strength from the ensemble'. This important aspect of the James-Stein estimators was illustrated in the articles of Efron and Morris (1975) with real data examples.

Stein (1960) also pointed out that the least squares estimators in a general linear regression model could be improved by certain 'shrinkage estimators' under squared error loss. A class of minimax estimators dominating the least squares estimators

under squared error loss was proposed in the articles of Sclove (1968) and Baranchik (1973). These articles, however, did not point out any empirical Bayes aspect of these minimax estimators.

Empirical Bayes motivation of James-Stein type shrinkage estimators dominating the least squares estimators in linear regression models is alluded to in Arnold's (1981) book (see Chapter 11). This phenomenon was brought out more explicitly in a recent article of Judge, Hill and Bock (1988).

The subject matter of this thesis stems more directly from an article of Ghosh, Saleh and Sen (1987). These authors proposed an empirical Bayes estimator of $\underline{\beta}_1$ in the general regression model

$$\underline{Y} = \underline{X}_1\underline{\beta}_1 + \underline{X}_2\underline{\beta}_2 + \underline{e}, \tag{1.1}$$

with $\underline{e} \sim N_n(\underline{0}, \sigma^2 I)$, where it was suspected that $\underline{\beta}_2 = \underline{0}$. The prior distribution of $\underline{\beta}^T = (\underline{\beta}_1^T, \underline{\beta}_2^T)$ was taken as $N_p(\underline{\nu}, \tau^2(\underline{X}^T\underline{X})^{-1})$ with $\underline{\nu}^T = (\underline{\nu}_1^T, \underline{0}^T)$, where $\underline{\nu}_1$ was a $p_1 \times 1$ vector with $p_1 < p$. The empirical Bayes estimator of $\underline{\beta}_1$,

$$\underline{e}_{EB}(\underline{Y}) = \hat{\underline{\beta}}_1 + (1 - \frac{(n-p)S_e^2(p_2-2)}{(n-p+2)(\tilde{\underline{\beta}}_2^T\underline{C}_{22.1}\tilde{\underline{\beta}}_2)})(\tilde{\underline{\beta}}_1 - \hat{\underline{\beta}}_1), \tag{1.2}$$

where $S_e^2 = \| \underline{Y} - \underline{X}\tilde{\underline{\beta}} \|^2/(n-p)$, served as compromise between $\tilde{\underline{\beta}}_1$, the least squares estimator(LSE) of $\underline{\beta}_1$ for the full model, and $\hat{\underline{\beta}}_1$, the reduced least squares estimator under the hypothesis $H_0 : \underline{\beta}_2 = \underline{0}$, and $\underline{C}_{22.1} = \underline{C}_{22} - \underline{C}_{21}\underline{C}_{11}^{-1}\underline{C}_{12}$ with $\underline{C}_{ij} = \underline{X}_i\underline{X}_j$ $i,j = 1,2$, and $\underline{X} = (\underline{X}_1\underline{X}_2)$. The compromise estimator leaned more towards $\hat{\underline{\beta}}_1$ when the data suggested that H_0 is true, and towards $\tilde{\underline{\beta}}_1$, otherwise. The optimality of $\underline{e}_{EB}(\underline{Y})$ was proved within the class of estimators of the form

$$\underline{\delta}_c(\underline{Y}) = \hat{\underline{\beta}}_1 + (1 - \frac{c(n-p)S_e^2}{(n-p+2)(\tilde{\underline{\beta}}_2^T\underline{C}_{22.1}\tilde{\underline{\beta}}_2)})(\tilde{\underline{\beta}}_1 - \hat{\underline{\beta}}_1), \tag{1.3}$$

where c is a constant, and $\underline{e}_{EB}(\underline{Y}) \equiv \underline{\delta}_{p_2-2}(\underline{Y})$. To be specific, under the loss

$$L(\underline{\beta}_1, \underline{a}) = (\underline{a} - \underline{\beta}_1)^T\underline{Q}(\underline{a} - \underline{\beta}_1)/\sigma^2, \tag{1.4}$$

where $\underline{Q}$ is a positive definite weight matrix, the Bayes risk of $\underline{\delta}_c$ under the aforementioned prior (to be denoted by ξ) is given by

$$r(\xi, \underline{\delta}_c) = (1-B)\, tr(\underline{Q}\underline{C}_{11.2}^{-1})$$
$$+B[tr(\underline{Q}\,\underline{C}_{11}^{-1}) + (1 - \frac{2c(n-p)}{p_2(n-p+2)} + \frac{c^2(n-p)}{p_2(n-p+2)(p_2-2)})tr(\underline{W}\,\underline{C}_{22.1}^{-1})] \quad (1.5)$$

where $B = \sigma^2/(\tau^2+\sigma^2)$, and $\underline{W} = \underline{C}_{21}\underline{C}_{11}^{-1}\underline{Q}\underline{C}_{11}^{-1}\underline{C}_{12}$. The Bayes risk is the expectation value of the loss (1.4) based on the marginal distribution of $\underline{Y}$. The risk $r(\xi, \underline{\delta}_c)$ is minimized at $c = p_2 - 2$, and $\underline{e}_{EB} = \underline{\delta}_{p_2-2}$. Note that $\underline{C}_{ij} = \underline{X}_i^T\underline{X}_j$ and $\underline{C}_{ii.j} = \underline{C}_{ii} - \underline{C}_{ij}\underline{C}_{jj}\underline{C}_{ji}$ $(i,j = 1,2)$.

Comparisons of the Bayes risk of $\underline{e}_{EB}$ with the ones for $\tilde{\underline{\beta}}_1$ and $\hat{\underline{\beta}}_1$ shows that

$$r(\xi, \tilde{\underline{\beta}}_1) - r(\xi, \underline{e}_{EB}) = B\frac{(n-p)(p_2-2)}{(n-p+2)p_2}\, tr(\underline{W}\,\underline{C}_{22.1}^{-1}) > 0, \quad (1.6)$$

so that the Bayes risk of $\underline{e}_{EB}$ is smaller than that of $\tilde{\underline{\beta}}_1$, and

$$r(\xi, \hat{\underline{\beta}}_1) - r(\xi, \underline{e}_{EB}) = \{(1-B)^2B^{-1} - B(1 - \frac{(n-p)(p_2-2)}{(n-p+2)p_2})\}\, tr(\underline{W}\,\underline{C}_{22.1}^{-1}) > 0, \quad (1.7)$$

so that the Bayes risk of $\underline{e}_{EB}$ is smaller than that of $\hat{\underline{\beta}}_1$ if and only if $(1-B)^2/B^2 > 1 - (n-p)(p_2-2)/(n-p+2)p_2$. The condition stated in (1.7) is expected since whenever the prior variance τ^2 is much smaller than the sampling variance σ^2, then B is close to 1, and in that case $\hat{\underline{\beta}}_1$ is expected to perform better than $\underline{e}_{EB}$ since the prior mean of $\underline{\beta}_2$ is set equal to zero.

Let $F = \tilde{\underline{\beta}}_2^T\underline{C}_{22.1}\tilde{\underline{\beta}}_2/(\frac{n-p}{n-p+2}S_e^2)$, which is a multiple of the usual F statistic of testing $H_0 : \underline{\beta}_2 = \underline{0}$, and

$$\tilde{\underline{\beta}}_1^G = \tilde{\underline{\beta}}_1 - \frac{\phi(F)}{F}(\tilde{\underline{\beta}}_1 - \hat{\underline{\beta}}_1), \quad (1.8)$$

which is the general class of estimators of $\underline{\beta}_1$. Then under the following three conditions

(i) $2\,ch_1(\underline{W}\,\underline{C}_{22.1}^{-1}) < tr(\underline{W}\,\underline{C}_{22.1}^{-1})$, where $ch_1(\underline{W}\,\underline{C}_{22.1}^{-1})$ denotes the largest eigenvalue of $(\underline{W}\,\underline{C}_{22.1}^{-1})$;

(ii) $0 < \phi(F) < 2\{tr(\underline{W}\,\underline{C}_{22.1}^{-1})/ch_1(\underline{W}\,\underline{C}_{22.1}^{-1}) - 2\}$;

(iii) $\phi(F)$ increasing monotonically in F,

it is found by Ghosh, Saleh, and Sen (1987) that the frequentist risk of $\tilde{\underline{\beta}}_{.1}^{G}$ is less than that of the minimax estimator of $\underline{\beta}_{.1}$, $\tilde{\underline{\beta}}_{.1}$, under the loss described in (1.4). Note that $\phi(F)$ is a general function of F and the frequentist risk of the estimators of $\underline{\beta}_{.1}$ is the expectation value of the loss (1.4) for fixed given value of $\underline{\beta}_{.1}$.

When $\phi(F) = p_2 - 2$,

$$0 < p_2 - 2 < 2\{tr(\underline{W}\,\underline{C}_{22.1}^{-1})/ch_1(\underline{W}\,\underline{C}_{22.1}^{-1}) - 2\}, \tag{1.9}$$

then if condition (i) holds, the empirical Bayes estimator will dominate $\tilde{\underline{\beta}}_{.1}$ and will be a minimax estimator of $\underline{\beta}_{.1}$.

When both the condition (i) and the condition described in (1.9) hold, the modified empirical Bayes estimators,

$$\delta_{EBM} = \tilde{\underline{\beta}}_{.1} - (1 - (1 - \frac{c}{F})g(F))(\tilde{\underline{\beta}}_{.1} - \hat{\underline{\beta}}_{.1}), \tag{1.10}$$

with $g(F) = I_{[F>d]}$ representing some positive constant d, and I denoting the usual indicator function, were shown to dominate the preliminary test estimators,

$$\delta_{PTE} = \tilde{\underline{\beta}}_{.1} - \{1 - g(F)\}(\tilde{\underline{\beta}}_{.1} - \hat{\underline{\beta}}_{.1}), \tag{1.11}$$

which were originally proposed by Bancroft (1944).

The prior distribution used by Ghosh, Saleh and Sen (1987) to motive their empirical Bayes estimators (see also Judge, Hill and Bock, 1988) has been called a g-prior by Zellner (1986). Zellner argued that g-priors could be viewed as reference informative prior distributions.

The conceptual idea behind the g-priors as described by Zellner (1986) is given as follows:

Before observing $\underline{Y}$, consider a 'conceptual' or 'imaginary' sample $\underline{Y}_0$, an $n \times 1$ vector, assumed to be generated by

$$\underline{Y}_0 = \underline{X}\,\underline{\beta} + \underline{u}, \tag{1.12}$$

where $\underline{u}$ is distributed as $N(\underline{0}, \sigma_0^2 \underline{V})$, and $\underline{X}$ is an $n \times p$ design matrix in the model of interest. The likelihood function for Eq.(1.12) is given by

$$l(\underline{\beta}, \sigma_0^2 | \underline{Y}_0) \propto \sigma_0^{-n} \exp\{-(\underline{Y}_0 - \underline{X}\underline{\beta})^T \underline{V}^{-1} (\underline{Y}_0 - \underline{X}\underline{\beta})/2\sigma_0^2\} \tag{1.13}$$

$$\propto \sigma_0^{-n} \exp\{-[(n-p)s_0^2 + (\underline{\beta} - \tilde{\underline{\beta}}_0)^T (\underline{X}^T \underline{V}^{-1} \underline{X})(\underline{\beta} - \tilde{\underline{\beta}}_0)]/2\sigma_0^2\},$$

where

$$\tilde{\underline{\beta}}_0 = (\underline{X}^T \underline{V}^{-1} \underline{X})^{-1} \underline{X}^T \underline{V}^{-1} \underline{Y}_0,$$

and

$$(n-p)s_0^2 = (\underline{Y}_0 - \underline{X}\tilde{\underline{\beta}}_0)^T \underline{V}^{-1} (\underline{Y}_0 - \underline{X}\tilde{\underline{\beta}}_0).$$

Little knowledge is known about the values of $\underline{\beta}$ and σ_0^2. Thus, one may use the diffuse prior (see, for example, Jeffreys, 1967)

$$p(\underline{\beta}, \sigma_0^2) \propto 1/\sigma_0. \tag{1.14}$$

This implies that the elements of $\underline{\beta}$ and $\log \sigma_0$ are independently and uniformly distributed. By combining Eqs.(1.13) and (1.14), the posterior distribution of $\underline{\beta}$ for given σ_0^2 will be normal with the mean $\tilde{\underline{\beta}}_0$ equal to $(\underline{X}^T \underline{V}^{-1} \underline{X})^{-1} \underline{X}^T \underline{V}^{-1} \underline{Y}_0$ and the variance-covariance matrix equal to $(\underline{X}^T \underline{V}^{-1} \underline{X})^{-1} \sigma_0^2$. Note that the posterior variance-covariance matrix is proportional to the variance-covariance matrix of the LSE $\tilde{\underline{\beta}} = (\underline{X}^T \underline{V}^{-1} \underline{X})^{-1} \underline{X}^T \underline{V}^{-1} \underline{Y}$ of $\underline{\beta}$. Thus, while carrying out the actual Bayesian analysis, the prior variance-covariance matrix is taken proportional to $(\underline{X}^T \underline{V}^{-1} \underline{X})^{-1}$.

The subject matter of this thesis first focuses on the response function $\underline{X}\tilde{\underline{\beta}}$ in univariate regression models $\underline{Y} = \underline{X}\underline{\beta} + \underline{e}$ (Khuri and Cornell, 1987). It then focuses on the multivariate general linear regression models which are often useful in growth and response curve studies. Empirical Bayes estimators for the multivariate general linear model where the individual regression coefficients $\underline{\beta}_k$'s are

random effect were proposed by Reinsel (1985) who placed particular emphasis on the estimators's mean squared error (MSE) matrix. His linear model for m distinct individuals was of the form

$$\underline{Y}_k = \underline{X}_k\underline{\beta}_k + \underline{e}_k,\ k = 1,\cdots,m. \tag{1.15}$$

The vectors $\underline{e}_k = (e_{1k},\cdots,e_{n_k k})'$ were assumed to be independently distributed as $N(\underline{0},\sigma^2\underline{I})$. It was assumed that the coefficients $\underline{\beta}_k$, with dimension $p\times 1$, were composed of both fixed effects that incorporate concomitant information and individual random effects such that

$$\underline{\beta}_k = \underline{B}\,\underline{a}_k + \underline{\lambda}_k,\ k = 1,\cdots,m, \tag{1.16}$$

where $\underline{A}' = (\underline{a}_1,\cdots,\underline{a}_m)$ is an $r\times m$ 'across individual' design matrix of constants of full rank $r < m$, whose k-th column $\underline{a}_k$ represents the values of the concomitant 'background' regressor variables associated with the k-th individual, and $\underline{B}$ is a $p\times r$ matrix of unknown parameters. The individual random effects $\underline{\lambda}_k$ are assumed to be distributed as $N(\underline{0},\underline{\Sigma}_\lambda)$, independent over individuals and independent of the errors $\underline{e}_k$. Eq. (1.16) can also be viewed as a prior distribution of $\underline{\beta}_k$. For the case that $\underline{X}_k = \underline{X}$ for all k and all parameters $\underline{B}$, σ^2, and $\underline{\Sigma}_\lambda$ are unknown, the proposed empirical Bayes estimator of $\underline{\beta}_k$ for the k-th individual dominated the usual least squares estimator of $\underline{\beta}_k$ with a smaller individual (unconditional) MSE matrix. According to Eqs (1.15) and (1.16), the Bayes estimator of $\underline{\beta}_k \equiv \underline{e}_B(\underline{Y}_k)$ is the posterior mean,

$$\underline{e}_B(\underline{Y}_k) = E(\underline{\beta}_k \mid \underline{Y}_k) = \tilde{\underline{\beta}}_k - \underline{V}_k(\underline{V}_k + \underline{\Sigma}_\lambda)^{-1}(\tilde{\underline{\beta}}_k - \underline{B}\underline{a}_k), \tag{1.17}$$

where $\tilde{\underline{\beta}}_k = (\underline{X}_k^T\underline{X}_k)^{-1}\underline{X}_k^T\underline{Y}_k$, the least squares estimator of $\underline{\beta}_k$, and $\underline{V}_k = \sigma^2(\underline{X}_k^T\underline{X}_k)^{-1}$. Hence, $\underline{e}_B(\underline{Y}_k)$ is the minimum MSE estimator of $\underline{\beta}_k$, with the unconditional MSE matrix given by

$$E[(\underline{\beta}_k - \underline{e}_B(\underline{Y}_k))(\underline{\beta}_k - \underline{e}_B(\underline{Y}_k))^T] = \underline{V}_k - \underline{V}_k(\underline{V}_k + \underline{\Sigma}_\lambda)^{-1}\underline{V}_k. \tag{1.18}$$

Under the models (1.15) and (1.16) with all $\underline{X}_k$ equal to $\underline{X}$, the empirical Bayes estimator has the following form:

$$\underline{\delta}_{EB}(\underline{Y}_k) = \tilde{\underline{\beta}}_k - \hat{\underline{V}}(\hat{\underline{V}} + \hat{\underline{\Sigma}}_\lambda)^{-1}(\tilde{\underline{\beta}}_k - \hat{\underline{B}}\underline{a}_k), \tag{1.19}$$

where $\tilde{\underline{\beta}}_k = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{Y}_k$, $\hat{\underline{V}} = (\frac{1}{m(n-p)})S_e^2(\underline{X}^T\underline{X})^{-1}$ with $S_e^2 = \sum_{k=1}^m \underline{Y}_k^T(\underline{I} - \underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}^T)\underline{Y}_k$, $\hat{\underline{B}} = \underline{\beta}^*\underline{A}(\underline{A}^T\underline{A})^{-1}$, $\underline{\beta}^* = (\tilde{\underline{\beta}}_1, \cdots, \tilde{\underline{\beta}}_m)^T$ and $\hat{\underline{\Sigma}}_\lambda = (m-r)^{-1}(\underline{\beta}^*)^T(\underline{I} - \underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T)\underline{\beta}^* - \hat{\underline{V}}$. Now consider estimators of individual $\underline{\beta}_k$ of the form:

$$\underline{\delta}_c(\underline{Y}_k) = \tilde{\underline{\beta}}_k - c\,\hat{\underline{V}}(\hat{\underline{V}} + \hat{\underline{\Sigma}}_\lambda)^{-1}(\tilde{\underline{\beta}}_k - \hat{\underline{B}}\underline{a}_k), \tag{1.20}$$

where c is an arbitrary constant. The unconditional MSE matrix of the estimator $\underline{\delta}_c(\underline{Y}_k)$ in Eq. (1.20) for any individual k is given by

$$\begin{aligned} E[(\underline{\delta}_c(\underline{Y}_k) - \underline{\beta}_k)(\underline{\delta}_c(Y_k) - \underline{\beta}_k)^T] &= \underline{V} + (1 - \underline{a}_k^T(\underline{A}^T\underline{A})^{-1}\underline{a}_k) \\ &\times\{\frac{c^2(m-r)(m(n-p)+2)}{(m-r-p-1)m(n-p)} - 2c\}\underline{V}(\underline{V} + \underline{\Sigma}_\lambda)^{-1}\underline{V}. \end{aligned} \tag{1.21}$$

The minimum value of (1.21) with respect to c occurs for $c = (m-r-p-1)m(n-p)/(m-r)(m(n-q)+2)$, with the minimum MSE matrix value of

$$\underline{V} - (1 - \underline{a}_k^T(\underline{A}^T\underline{A})^{-1}\underline{a}_k)\{\frac{(m-r-p-1)m(n-p)}{(m-r)(m(n-p)+2)}\}\underline{V}(\underline{V} + \underline{\Sigma}_\lambda)^{-1}\underline{V} \tag{1.22}$$

Note from (1.22) that the MSE matrix of the estimator (1.20) with the optimal value of c is smaller than that of the individual least squares estimator $\tilde{\underline{\beta}}_k$ for every individual k, provided the $m > r + p + 1$.

Reinsel (1985) also investigates the MSE matrix properties of the estimators $\underline{\delta}_c(\underline{Y}_k)$ in the frequentist sense (i.e. fixed given values of $\underline{\beta}_1, \cdots, \underline{\beta}_m$). The estimators $\underline{\delta}_c(\underline{Y}_k)$ may still be appropriate in this setting as long as the concomitant variables $\underline{a}_k$ are capable of explaining a significant portion of the variation among the values $\underline{\beta}_k$. Now, consider only the case of equal individual design matrices $\underline{X}_k = \underline{X}$. Let $E_{\underline{\beta}}$ denote expectation conditional on the values of $\underline{\beta} = (\underline{\beta}_1, \cdots, \underline{\beta}_m)$. Then

$$\frac{1}{m}\sum_{k=1}^{m} E_{\underline{\beta}}[(\underline{\delta}_c(\underline{Y}_k) - \underline{\beta}_k)(\underline{\delta}_c(\underline{Y}_k) - \underline{\beta}_k)^T]$$

$$= \underline{V} + \{c^2\frac{(m-r)^2(m(n-p)+2)}{m(m(n-p))} - \frac{2c(m-r)}{m}(m-r-p-1)\}$$

$$\times \underline{V}E_{\underline{\beta}}(\underline{S}_\lambda^{-1})\underline{V}, \tag{1.23}$$

where $\underline{S}_\lambda = \underline{\beta}^{*T}(\underline{I} - \underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T)\underline{\beta}^*$. As with the unconditional expression ,the preceding MSE matrix expression is minimized by the choice $c = (m-r-p-1)m(n-p)/(m-r)(m(n-p)+2)$, with the minimum value for Eq. (1.23) equal to

$$\underline{V} - [(m-r-p-1)^2 m(n-p)/m(m(n-p)+2)]\underline{V}E_{\underline{\beta}}(\underline{S}_\lambda^{-1})\underline{V}. \tag{1.24}$$

Thus the estimator $\underline{\delta}_c(\underline{Y}_k)$ dominates the individual $\underline{\tilde{\beta}}_k$ in terms of the average MSE matrix conditionally for any fixed values of $\underline{\beta}_1, \cdots, \underline{\beta}_m$, provided $m > r + p + 1$ and $0 < c < 2(m-r-p-1)m(n-p)/\{(m-r)(m(n-p)+2)\}$. Reinsel (1985) also investigated this procedure in certain mixed models and in more general multiparameter situations such as the multivariate one-way random effects model.

1.2 Outline of Work

In response surface analysis, estimation of response functions is often of interest. For doing the estimation, there may be a choice between the following two nested regression models

$$\underline{Y} = \underline{X}_1\underline{\beta}_1 + \underline{X}_2\underline{\beta}_2 + \underline{e} \tag{1.25}$$

and

$$\underline{Y} = \underline{X}_1\underline{\beta}_1 + \underline{e}, \tag{1.26}$$

where $\underline{e} \sim N(\underline{0}, \underline{V})$. For example, in factorial experiments, one may want to select one of the two models, the first involving the main effects as well as the interaction effects, and the second involving only the main effects. For a given set of data, using the least squares method, the full model always produces a better fit than the reduced model. However, this may not be the case for estimating $\underline{X}\underline{\beta}$ ($\underline{X} = (\underline{X}_1, \underline{X}_2)$, $\underline{\beta}^T = (\underline{\beta}_1^T, \underline{\beta}_2^T)$). More specially, assuming $\underline{X}^T\underline{V}^{-1}\underline{X}$ to be nonsingular, if

$\tilde{\underline{\beta}}$ is the LSE of $\underline{\beta}$ based on the full model, and the $\hat{\underline{\beta}}_1$ is the LSE of $\underline{\beta}_1$ for the reduced model, then $\|\underline{Y} - \underline{X}\tilde{\underline{\beta}}\|^2 \leq \|\underline{Y} - \underline{X}_1\hat{\underline{\beta}}_1\|^2$. This does not necessarily imply that $\|\underline{X}\tilde{\underline{\beta}} - \underline{X}\underline{\beta}\|^2 \leq \|\underline{X}_1\hat{\underline{\beta}}_1 - \underline{X}\underline{\beta}\|^2$, or that $E\|\underline{X}\tilde{\underline{\beta}} - \underline{X}\underline{\beta}\|^2 \leq E\|\underline{X}_1\hat{\underline{\beta}}_1 - \underline{X}\underline{\beta}\|^2$ for all $\underline{\beta}_1$ and $\underline{\beta}_2$ under the usual normal model.

One classical way to resolve this problem is to test the null hypothesis $H_0 : \underline{\beta}_2 = \underline{0}$, and then use the full model if the null hypothesis is rejected at a desired level of significance, and use the reduced model if the opposite is true. The above procedure suffers from the drawback that there is no general way of incorporating the degree of evidence for or against the null hypothesis in order to arrive at the estimator. For instance, if observed F values are very close to a critical value (at a chosen level of signicance), then there may be no persuasive reason for choosing one model in preference to the other.

In Chapter 2, empirical (EB) and hierarchical Bayes (HB) estimators of $\underline{X}\underline{\beta}$ are developed. Instead of estimating the unknown parameters of the Bayes estimator from the data, a hierarchical Bayes method is that one put a prior distribution, often improper, on the unknown hyperparameters and arrive at the posterior distribution of $\underline{\beta}$ given $\underline{Y} = \underline{y}$. Zellner's (1986) g-prior is utilized to obtain a Bayes estimator of $\underline{X}\underline{\beta}$. The structure of the prior mean is motivated from the null hypothesis $H_0 : \underline{\beta}_2 = \underline{0}$. Empirical Bayes ($EB$), positive part EB, and modified EB estimators are introduced in Section 2.2. Under a general quadratic loss, the performance of the empirical Bayes estimators of $\underline{X}\underline{\beta}$ is evaluated using both the frequentist and the Bayes criteria. Under certain conditions, it is shown that the empirical Bayes estimators dominate the minimax estimator, the LSE of $\underline{X}\underline{\beta}$ in a frequentist sense. Thus a group of minimax estimators of $\underline{X}\underline{\beta}$ is provided. The empirical Bayes estimators are also shown to have smaller Bayes risks than the LSE for the full model. It should be noted though that like the empirical Bayes estimators, the preliminary test estimators (PTE) serve as compromises between

the LSE's based on the full and the reduced models. The PTE is dominated by the corresponding modified empirical Bayes estimator since the former does not take account of the degree of evidence for or against the null hypothesis $H_0 : \underline{\beta}_2 = \underline{0}$. Several theorems which describe the frequentist as well as the Bayesian properties of the proposed estimators are given in Section 2.2. In Section 2.3 some important HB estimators are introduced and the connection between the HB and EB estimators are discussed. An advantage of the HB procedure over the EB procedure is that the former can be used to construct credible sets for $\underline{X}\underline{\beta}$. Several numerical examples including real data and simulated data are provided in Section 2.4 to illustrate the applicability of the methods proposed in Sections 2.2 and 2.3.

This research further studies the multivariate linear regression model which is often useful in growth and response curve analyses. (see e.g. Reinsel, 1985). Suppose there is a choice between the following multivariate regression models.

$$\underline{Y}_k = \underline{X}_{k1}\underline{\beta}_{k1} + \underline{X}_{k2}\underline{\beta}_{k2} + \underline{e}_k \tag{1.27}$$

$$\underline{Y}_k = \underline{X}_{k1}\underline{\beta}_{k1} + \underline{e}_k \tag{1.28}$$

where $\underline{e}_k \sim N(\underline{0}, \underline{V}_k)$ for $k = 1, \cdots, m$. the model given in (1.28) is based on the guess that $\underline{\beta}_{k2} = \underline{0}$ for all $k = 1, \cdots, m$.

Instead, one can also compare the models (1.27) and

$$\underline{Y}_k = \underline{X}_{k1}\underline{\beta}_{1.} + \underline{e}_k, \ k = 1, \cdots, m. \tag{1.29}$$

The model (1.29) is not only based on the guess that $\underline{\beta}_{k2} = \underline{0} \ \forall \ k = 1, \cdots, m$ but also on the guess that $\underline{\beta}_{11}, \cdots, \underline{\beta}_{m1}$ have a common value $\underline{\beta}_{1.}$.

Furthermore, for the problem of best fitting of a given set of data by using the least squares method, the model in (1.27) behaves better than the reduced models given in (1.28) and (1.29). However, when the estimators of $(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m)$ are compared on the basis of general quadratic loss,

$$\mathrm{L}(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{b}_1, \cdots, \underline{b}_m) = \sum_{k=1}^{m} (\underline{b}_k - \underline{X}_k\underline{\beta}_k)^T \underline{Q}_k (\underline{b}_k - \underline{X}_k\underline{\beta}_k), \tag{1.30}$$

where $\underline{Q}_k$ is a known positive definite matrix, the LSE's of $\underline{X}_k\underline{\beta}_k$ based on the model in (1.27) may not keep their superiority.

Chapter 3 introduces the empirical Bayes estimators solving the choice between the models in (1.27) and (1.29). In order to motivate the Bayes estimators of $(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m)$, Zellner's (1986) g-prior is used once again.

For the choice between the models given in (1.27) and (1.28), the analysis follows closely the ones described in Chapter 2. However, if more information is available about the parameters in the model (1.28), Reinsel's (1985) prior distribution of $(\underline{\beta}_1, \cdots, \underline{\beta}_m)$ may be adopted and modified in finding the Bayes estimators. This is what is used in Chapter 4 to motivate the empirical Bayes estimators assuming $\underline{X}_1 = \cdots = \underline{X}_m = \underline{X}$. These empirical Bayes estimators serve as a compromise between the LSE's in the full and the reduced models given in (1.27) and (1.28) where $\underline{X}_1 = \cdots = \underline{X}_m = \underline{X}$.

The outlines of Chapter 3 and Chapter 4 are similar. In Section 3.2 and Section 4.2, empirical Bayes estimators and positive part empirical Bayes estimators are introduced, respectively. These estimators are then evaluated on the basis of the frequentist as well as the Bayesian risks. It is shown that there exists empirical Bayes estimators of $(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m)$ or $(\underline{X}\underline{\beta}_1, \cdots, \underline{X}\underline{\beta}_m)$ which always dominates the LSE of the model in (1.27) under squared error loss. The corresponding robustness and minimax estimation properties are also introduced. The preliminary test estimators serve as compromises between the LSE's based on the model in (1.27) and the model in (1.29) (Chapter 3), or between the LSE of the model in (1.27) and the LSE in the model (1.28) (Chapter 4). Also, it is shown that, there are modified empirical Bayes estimators which dominate the preliminary test estimators.

In both Section 3.3 and 4.3 several estimators of $\underline{\beta}_{k1}$ are proposed which shrink the unrestricted LSE, $\underline{\tilde{\beta}}_{k1}$, to the restricted LSE, $\underline{\hat{\beta}}_{k1}$ or $\underline{\hat{\beta}}_{1\cdot}$, (Chapter 3) or to a weighted average of the LSE's $\underline{\hat{\beta}}_{k1}$'s (Chapter 4). Under squared error loss,

frequentist risks as well as Bayes risks of these estimators are compared. Numerical simulations illustrating the applicability of the estimators introduced in Sections 3.2 and 4.2 are then given in Sections 3.4 and 4.4 respectively.

Finally, a summary of the present work and some suggestions for future study are given in Chapter 5.

CHAPTER 2
EMPIRICAL AND HIERARCHICAL BAYES ESTIMATION OF THE RESPONSE FUNCTION

2.1 Introduction

In the analysis of data from designed experiments, especially in response surface analysis, often the problem is estimation of the response function (see, e.g., Chaper 6 of Khuri and Cornell, 1987). For doing so, there may be a choice between the two nested regression models

$$\begin{aligned} \underline{Y} &= \underline{X}\,\underline{\beta} + \underline{e} \\ &= \underline{X}_1\,\underline{\beta}_1 + \underline{X}_2\,\underline{\beta}_2 + \underline{e} \end{aligned} \tag{2.1}$$

and

$$\underline{Y} = \underline{X}_1\,\underline{\beta}_1 + \underline{e}, \tag{2.2}$$

where $\underline{Y}$ is an $n \times 1$ vector of observations, $\underline{X} = (\underline{X}_1, \underline{X}_2)$ is an $n \times p$ design matrix which is of rank p, the dimension of $\underline{X}_1$ and $\underline{X}_2$ are $n \times p_1$ and $n \times p_2$, respectively, with $p = p_1 + p_2$, $\underline{\beta}^T = (\underline{\beta}_1^T, \underline{\beta}_2^T)$ is a $p \times 1$ vector of regression parameters with unknown values, and $\underline{e}$, is an $n \times 1$ vector of error terms. It is assumed that $\underline{e}$ is distributed as $N(\underline{0}, \sigma^2 \underline{V})$, where $\underline{V}$ is a positive definite (p.d.) matrix. For example, in factorial experiments, one may want to select one of the two models, the first involving the main effects as well the interactions, and the second involving only the main effects. One classical way to settle this problem is to test the null hypothesis $H_0 : \underline{\beta}_2 = \underline{0}$, and then use model (2.1) if the null hypothesis is rejected at a desired level of significance, and use model (2.2), otherwise. The above procedure suffers from the drawback that there is no general way of incorporating the degree of

evidence for or against the null hypothesis in order to arrive at the estimator. For example, if the observed F value is very close to the critical value (at a chosen level of significance), then there seems to be no convincing reason for choosing one model in preference to the other.

For the problem of best fitting of a given set of data by the least squares methods, the full model always beats the reduced model. However, the same need not necessarily hold for estimating $\underline{X}\underline{\beta}$. More specifically, let $\tilde{\underline{\beta}} = (\underline{X}^T \underline{V}^{-1} \underline{X})^{-1} \underline{X}^T \underline{V}^{-1} \underline{Y}$ denote the least squares estimator (LSE) of $\underline{\beta}$ based on model (2.1), and $\hat{\underline{\beta}}_1 = (\underline{X}_1^T \underline{V}^{-1} \underline{X}_1)^{-1} \underline{X}_1 \underline{V}^{-1} \underline{Y}$ denote the LSE of $\underline{\beta}_1$ for model (2.2), then $(\underline{Y} - \underline{X}\tilde{\underline{\beta}})^T \underline{V}^{-1} (\underline{Y} - \underline{X}\tilde{\underline{\beta}}) \leq (\underline{Y} - \underline{X}_1 \hat{\underline{\beta}}_1)^T \underline{V}^{-1} (\underline{Y} - \underline{X}_1 \hat{\underline{\beta}}_1)$, but that does not necessarily imply that $(\underline{X}\underline{\beta} - \underline{X}\tilde{\underline{\beta}})^T \underline{V}^{-1} (\underline{X}\underline{\beta} - \underline{X}\tilde{\underline{\beta}}) \leq (\underline{X}\underline{\beta} - \underline{X}_1 \hat{\underline{\beta}}_1)^T \underline{V}^{-1} (\underline{X}\underline{\beta} - \underline{X}_1 \hat{\underline{\beta}}_1)$ or even $E(\underline{X}\underline{\beta} - \underline{X}\tilde{\underline{\beta}})^T \underline{V}^{-1} (\underline{X}\underline{\beta} - \underline{X}\tilde{\underline{\beta}}) \leq E(\underline{X}\underline{\beta} - \underline{X}_1 \hat{\underline{\beta}}_1)^T \underline{V}^{-1} (\underline{X}\underline{\beta} - \underline{X}_1 \hat{\underline{\beta}}_1)$ for all $\underline{\beta}_1, \underline{\beta}_2$.

In this chapter, empirical (EB) and hierarchical Bayes (HB) estimators of $\underline{X}\underline{\beta}$ are proposed. The EB and the HB estimators serve as weighted average of the LSE's of $\underline{X}\underline{\beta}$ based on the full model (2.1) and the reduced model (2.2). These weights are data dependent, and are such that the proposed estimators lean more towards $\underline{X}\tilde{\underline{\beta}}$ if the F ratio for testing $H_0 : \underline{\beta}_2 = \underline{0}$ is large, and to $\underline{X}_1 \hat{\underline{\beta}}_1$ when the opposite is true. In fact, it is shown in Section 2.2 that if there is sufficient strong evidence for or against the null hypothesis, then some of the proposed EB estimators may indeed equal $\underline{X}\tilde{\underline{\beta}}$ or $\underline{X}_1 \hat{\underline{\beta}}_1$, as the case may be.

Without shrinking the unrestricted least squares estimator $\underline{X}_1 \tilde{\underline{\beta}}_1$ to the restricted least squares estimator $\underline{X}_1 \hat{\underline{\beta}}_1$ but shrinking $\underline{X}_2 \tilde{\underline{\beta}}_2$ towards zero, another estimator was proposed by Sclove (1968). Sclove's estimator serves as a compromise between $\underline{X}_1 \tilde{\underline{\beta}}_1$ and $\underline{X}\tilde{\underline{\beta}}$, and is closer to $\underline{X}_1 \tilde{\underline{\beta}}_1$ if $H_0 : \underline{\beta}_2 = \underline{0}$ is true, and close $\underline{X}\tilde{\underline{\beta}}$ otherwise. It will be shown in Section 2.2 that the EB estimator enjoys both Bayesian and frequentist risk superiority over the estimator developed by Sclove.

The result is expected, since $\underline{X}_1\hat{\underline{\beta}}_1$ is a more appropriate estimator of $\underline{X}\underline{\beta}$ than $\underline{X}_1\tilde{\underline{\beta}}_1$ when $H_0 : \underline{\beta}_2 = \underline{0}$ holds.

The preliminary test estimators (PTE's) provide yet another compromise between the LSE's based on full and reduced model. In Section 2.2, it will be seen that for every PTE of $\underline{X}\underline{\beta}$, there is a corresponding modified EB estimator which dominates the PTE. This consequence is not surprised, since the degree of evidence for or against the null hypothesis $H_0 : \underline{\beta}_2 = \underline{0}$ is not considered in the PTE's.

The EB estimators and their positive part versions are introduced in Section 2.2. Several theorems are given which describe the frequentist as well as the Bayesian properties of the proposed estimators. Specifically, it is shown that there are EB estimators of $\underline{X}\underline{\beta}$ which always dominate $\underline{X}\tilde{\underline{\beta}}$ under squared error loss. In Section 2.3, the HB estimators is introduced which also serve as a compromise between the LSE's based on full and reduced models. The link between the HB and EB estimators is also pointed out. One advantage of a HB procedure over an EB procedure is that the former can be also for the construction of credible sets for $\underline{X}\underline{\beta}$. Some real data examples are given in Section 2.4 to illustrate the estimation methods discussed in Section 2.2 and 2.3. Also, in two simple response surface examples, the different estimators of $\underline{X}\underline{\beta}$ are compared according to their simulated frequentist and Bayes risks.

Some of the results of Section 2.2 are related to the work of Ghosh, Saleh and Sen (1987), where a class of estimators of $\underline{\beta}_1$ is proposed compromising the LSE's based on the full and reduced models. They adopt entirely a classical EB approach whereas the present thesis introduces and relates both the EB and HB approaches.

2.2 The Empirical Bayes Approach

In order to motive the EB approach, a Bayes estimator of $\underline{X}\underline{\beta}$ has to be found. A general Bayes model for the present situation is

$$\underline{Y} \mid \underline{\beta} \sim N_n(\underline{X}\underline{\beta}, \sigma^2\underline{V}) \tag{2.3}$$

and

$$\underline{\beta} \sim N_p(\underline{\nu}, \underline{\Sigma}), \tag{2.4}$$

where $\underline{\nu}^T = (\underline{\nu}_1^T, \underline{0}^T)$ with $p_1 + p_2 = p$, $\underline{V}$ is known positive definite (p.d.) and $\underline{\Sigma}$ is p.d. as well. The structure of the prior is motivated from the null hypothesis $H_0 : \underline{\beta}_2 = \underline{0}$ mentioned in Section 2.1. Instead of considering the above general prior, following Ghosh, Saleh and Sen (1987), Zellner's (1986) g-prior is considered, where $\underline{\Sigma} = \tau^2(\underline{X}^T \underline{V}^{-1} \underline{X})^{-1}$. The g-prior has been justified by Zellner (1986) using Muth's (1961) rational expectations hypothesis, and the conceptual idea behind the use of g-prior is introduced in Section 1.1.

With the above g-prior, the posterior distribution of $\underline{\beta} \mid \underline{Y}$ is

$$N_p[\underline{\nu} + (1 - B)(\underline{\tilde{\beta}} - \underline{\nu}), \sigma^2(1 - B)(\underline{X}^T \underline{V}^{-1} \underline{X})^{-1}]$$

where

$$\underline{\tilde{\beta}} = (\underline{X}^T \underline{V}^{-1} \underline{X})^{-1} \underline{X}^T \underline{V}^{-1} \underline{Y}$$

is the $BLUE$ of $\underline{\beta}$, and $B = \sigma^2/(\sigma^2 + \tau^2)$. Writing $\underline{\tilde{\beta}}^T = (\underline{\tilde{\beta}}_1^T, \underline{\tilde{\beta}}_2^T)$, the Bayes estimator of $\underline{X}\underline{\beta}$ under any quadratic loss

$$\mathrm{L}(\underline{X}\underline{\beta}, \underline{a}) = (\underline{a} - \underline{X}\underline{\beta})^T \underline{Q} (\underline{a} - \underline{X}\underline{\beta}), \tag{2.5}$$

$\underline{Q}$ being a known p.d. matrix, is given by

$$\underline{e}_B(\underline{Y}) = E(\underline{X}\underline{\beta} \mid \underline{Y}) = \underline{X}_1[\underline{\nu}_1 + (1 - B)(\underline{\tilde{\beta}}_1 - \underline{\nu}_1)] + (1 - B)\underline{X}_2 \underline{\tilde{\beta}}_2. \tag{2.6}$$

Suppose the parameters $\underline{\nu}_1$, σ^2 and τ^2 are unknown and need to be estimated from the data. There are two ways to handle this situation. One is the EB method, which will be discussed here, where $\underline{\nu}_1$, σ^2 and τ^2 are estimated in some classical way such as the MLE's, $UMVUE$'s, best invariant estimators etc. based on the marginal distribution of $\underline{Y}$. The other is the so-called HB method, discussed in Section 2.3, where some prior distribution, quite often improper, on the unknown

parameters is provided, and inferences is made based on the resulting posterior distribution.

In order to generate the EB estimator, first note that marginally

$$\underline{Y} \sim N_n(\underline{X}_1 \underline{\nu}_1, \sigma^2 \underline{V} + \tau^2 \underline{P}_{\underline{X}}), \tag{2.7}$$

where $\underline{P}_{\underline{X}} = \underline{X}(\underline{X}^T \underline{V}^{-1} \underline{X})^{-1} \underline{X}^T$. Using Exercise 2.9, p. 33 of Rao (1973), one can write

$$(\sigma^2 \underline{V} + \tau^2 \underline{P}_{\underline{X}})^{-1} = \sigma^{-2}[\underline{V}^{-1} - (1-B)\underline{V}^{-1} \underline{P}_{\underline{X}} \underline{V}^{-1}], \tag{2.8}$$

where one may recall that $B = \sigma^2/(\sigma^2 + \tau^2)$. The $BLUE$ of $\underline{\beta}_1$ for the reduced model $\underline{Y} = \underline{X}_1 \underline{\beta}_1 + \underline{e}$ is $\hat{\underline{\beta}}_1 = (\underline{X}_1^T \underline{V}^{-1} \underline{X}_1)^{-1} \underline{X}_1^T \underline{V}^{-1} \underline{Y}$, and it is easy to see that

$$\tilde{\underline{\beta}}_1 = \hat{\underline{\beta}}_1 - \underline{C}_{11}^{-1} \underline{C}_{12} \tilde{\underline{\beta}}_2, \tag{2.9}$$

where $\underline{C}_{ij} = \underline{X}_i^T \underline{V}^{-1} \underline{X}_j$ for i=1,2 and j=1,2.

Using Eqs.(2.8) and (2.9), $\underline{P}_{\underline{X}} \underline{V}^{-1} \underline{X}_i = \underline{X}_i$ for i=1,2, and the idempotency of $\underline{V}^{-1} \underline{P}_{\underline{X}}$, after some algebra one gets

$$\begin{aligned}
& (\underline{Y} - \underline{X}_1 \underline{\nu}_1)^T (\sigma^2 \underline{V} + \tau^2 \underline{P}_{\underline{X}})^{-1} (\underline{Y} - \underline{X}_1 \underline{\nu}_1) \\
= \; & \sigma^{-2} (\underline{Y} - \underline{X} \tilde{\underline{\beta}} + \underline{X} \tilde{\underline{\beta}} - \underline{X}_1 \hat{\underline{\beta}}_1 + \underline{X}_1 \hat{\underline{\beta}}_1 - \underline{X}_1 \underline{\nu}_1)^T \\
& (\underline{V}^{-1} - (1-B) \underline{V}^{-1} \underline{P}_{\underline{X}} \underline{V}^{-1}) (\underline{Y} - \underline{X} \tilde{\underline{\beta}} + \underline{X} \tilde{\underline{\beta}} - \underline{X}_1 \hat{\underline{\beta}}_1 + \underline{X}_1 \hat{\underline{\beta}}_1 - \underline{X}_1 \underline{\nu}_1) \\
= \; & B \sigma^{-2} [SSE + SSH + (\hat{\underline{\beta}}_1 - \underline{\nu}_1)^T \underline{C}_{11} (\hat{\underline{\beta}}_1 - \underline{\nu}_1)],
\end{aligned} \tag{2.10}$$

where $SSE = (\underline{Y} - \underline{X} \tilde{\underline{\beta}})^T \underline{V}^{-1} (\underline{Y} - \underline{X} \tilde{\underline{\beta}})$ is the usual error sum of squares, and $SSH = \tilde{\underline{\beta}}_2^T \underline{C}_{22.1} \tilde{\underline{\beta}}_2$ with $\underline{C}_{22.1} = \underline{C}_{22} - \underline{C}_{21} \underline{C}_{11}^{-1} \underline{C}_{12}$ is the sum of squares due to the hypothesis $H_0 : \underline{\beta}_2 = \underline{0}$. As in Theorem 2 of Ghosh, Saleh and Sen (1987), it is now easy to see that:

$(\hat{\underline{\beta}}_1, SSH, SSE)$ is complete sufficient for $\underline{\nu}_1, \tau^2$ and σ^2.

Since $SSH \sim (\tau^2 + \sigma^2) \chi^2_{p_2}$, the $UMVUE$ of $\sigma^{-2} B = (\tau^2 + \sigma^2)^{-1}$ is $(p_2 - 2)/SSH$ for $p_2 \geq 3$. The best scale invariant estimate of σ^2 is $SSE/(n - p + 2)$, where

$p = p_1 + p_2$, and $UMVUE$ of ν_1 is $\hat{\underline{\beta}}_1$. Plugging all these estimators for the unknown parameters in (2.6), the EB estimator of $\underline{X}\underline{\beta}$ is given by

$$\underline{e}_{EB}(\underline{Y}) = \underline{X}_1[\hat{\underline{\beta}}_1 + (1 - \frac{p_2 - 2}{F})(\tilde{\underline{\beta}}_1 - \hat{\underline{\beta}}_1)] + (1 - \frac{p_2 - 2}{F})\underline{X}_2\tilde{\underline{\beta}}_2, \tag{2.11}$$

where $F = (SSH/SSE)(n - p + 2)$, a constant multiple of the usual F ratio for testing $H_0 : \underline{\beta}_2 = \underline{0}$.

Due to the nature of the above method of estimation, B is estimated by the quantity $(p_2 - 2)/F$ which can exceed 1. For practical purposes, instead of using the EB estimator in (2.11), the positive part EB estimator

$$\underline{e}^+_{EB}(\underline{Y}) = \underline{X}_1[\hat{\underline{\beta}}_1 + (1 - \frac{p_2 - 2}{F})^+(\tilde{\underline{\beta}}_1 - \hat{\underline{\beta}}_1)] + (1 - \frac{p_2 - 2}{F})^+\underline{X}_2\tilde{\underline{\beta}}_2, \tag{2.12}$$

where $a^+ = \max(a, 0)$, should be used.

Note from (2.12) that for very large F values indicating substantial departure from $H_0 : \underline{\beta}_2 = \underline{0}$, $\underline{e}^+_{EB}$ is very close to $\underline{X}\tilde{\underline{\beta}}$, whereas for very small F values indicating enough support for H_0, $\underline{e}^+_{EB}$ is very close to $\underline{X}_1\hat{\underline{\beta}}_1$. When there is no clear-cut decision for or against H_0, $\underline{e}^+_{EB}$ is in some sense a weighted average of $\underline{X}\tilde{\underline{\beta}}$ and $\underline{X}_1\hat{\underline{\beta}}_1$, and the weights being adaptively determined by the data.

Moreover, if the design matrix $\underline{X}$ is such that $\underline{C}_{12} = \underline{0}$, then from (2.9), $\tilde{\underline{\beta}}_1 = \hat{\underline{\beta}}_1$. $\underline{e}_{EB}$ definded in (2.11) simplifies to $\underline{e}_{EB}(\underline{Y}) = \underline{X}_1\hat{\underline{\beta}}_1 + (1 - (p_2 - 2)/F)\underline{X}_2\tilde{\underline{\beta}}_2$. Correspondingly, $\underline{e}^+_{EB}(\underline{Y})$ also simplifies.

The estimators $\underline{e}_{EB}$ and $\underline{e}^+_{EB}(\underline{Y})$ are evaluated according to the criteria of frequentist and Bayes risks. First, the frequentist criterion is discussed.

It is important to note that $\underline{e}_{EB}$ is a member of a class of estimators of the form

$$\underline{e}_\phi(\underline{Y}) = \underline{X}_1[\hat{\underline{\beta}}_1 + (1 - \frac{\phi(F)}{F})(\tilde{\underline{\beta}}_1 - \hat{\underline{\beta}}_1)] + (1 - \frac{\phi(F)}{F})\underline{X}_2\tilde{\underline{\beta}}_2. \tag{2.13}$$

The following theorem provides the risk expression for $\underline{e}_\phi$ under the loss (2.5). The risk is denoted as $R(\underline{X}\underline{\beta}, \underline{e}_\phi)$.

Theorem 2.2.1. If ϕ is a differentiable function, then under the loss (2.5),

$$\begin{aligned} R(\underline{X}\,\underline{\beta}, \underline{e}_\phi) &= R(\underline{X}\,\underline{\beta}, \underline{X}\,\tilde{\underline{\beta}}) - 4\sigma^2 E_{\underline{\beta},\sigma^2}[\phi'(F)\tilde{\underline{\beta}}_2^T \underline{G}\,\tilde{\underline{\beta}}_2/SSH] \\ &\quad -2\sigma^2 E_{\underline{\beta},\sigma^2}[\frac{\phi(F)}{F}\{tr(\underline{G}\,\underline{C}_{22.1}^{-1}) - 2(\tilde{\underline{\beta}}_2^T \underline{G}\,\tilde{\underline{\beta}}_2)/SSH\}] \\ &\quad +E_{\underline{\beta},\sigma^2}[\{\phi^2(F)/F^2\}\tilde{\underline{\beta}}_2^T \underline{G}\,\tilde{\underline{\beta}}_2], \end{aligned} \tag{2.14}$$

where $\underline{G} = (\underline{X}_2 - \underline{X}_1\,\underline{C}_{11}^{-1}\,\underline{C}_{12})^T \underline{Q}(\underline{X}_2 - \underline{X}_1\,\underline{C}_{11}^{-1}\,\underline{C}_{12})$.

Proof of Theorem 2.2.1. Now, using

$$\tilde{\underline{\beta}}_1 = \hat{\underline{\beta}}_1 - \underline{C}_{11}^{-1}\,\underline{C}_{12}\,\tilde{\underline{\beta}}_2, \tag{2.15}$$

$\underline{e}_\phi(\underline{Y})$ can be written as

$$\underline{e}_\phi(\underline{Y}) = \underline{X}_1\,\hat{\underline{\beta}}_1 + (1-\phi(F)/F)(\underline{X}_2 - \underline{X}_1\,\underline{C}_1^{-1}\underline{C}_{12})\tilde{\underline{\beta}}_2. \tag{2.16}$$

Since $E_{\underline{\beta},\sigma^2}(\hat{\underline{\beta}}_1) = \underline{\beta}_1 + \underline{C}_{11}^{-1}\underline{C}_{12}\underline{\beta}_2$ and $E_{\underline{\beta},\sigma^2}(\tilde{\underline{\beta}}_2) = \underline{\beta}_2$, and one writes

$$\underline{X}\,\underline{\beta} = \underline{X}_1\,\underline{\beta}_1 + \underline{X}_2\,\underline{\beta}_2 = \underline{X}_1\,E_{\underline{\beta},\sigma^2}(\hat{\underline{\beta}}_1) + (\underline{X}_2 - \underline{X}_1\underline{C}_{11}^{-1}\underline{C}_{12})\underline{\beta}_2. \tag{2.17}$$

Now, using the independence of $\hat{\underline{\beta}}_1$ and $\tilde{\underline{\beta}}_2$, it follows from (2.16) and (2.17) that under the loss (2.5),

$$\begin{aligned} & R(\underline{X}\,\underline{\beta}, \underline{e}_\phi) \\ &= E_{\underline{\beta},\sigma^2}\{(\hat{\underline{\beta}}_1 - E_{\underline{\beta},\sigma^2}(\hat{\underline{\beta}}_1))^T(\underline{X}_1^T\underline{Q}\underline{X}_1)(\hat{\underline{\beta}}_1 - E_{\underline{\beta},\sigma^2}(\hat{\underline{\beta}}_1)) \\ &+ (\tilde{\underline{\beta}}_2 - \underline{\beta}_2 - \frac{\phi(F)}{F}\tilde{\underline{\beta}}_2)^T\underline{G}(\tilde{\underline{\beta}}_2 - \underline{\beta}_2 - \frac{\phi(F)}{F}\tilde{\underline{\beta}}_2)\} \\ &= tr\,\{(\underline{X}_1^T\underline{Q}\underline{X}_1)V(\hat{\underline{\beta}}_1)\} + E_{\underline{\beta},\sigma^2}\{(\tilde{\underline{\beta}}_2 - \underline{\beta}_2)^T\underline{G}(\tilde{\underline{\beta}}_2 - \underline{\beta}_2)\} \\ &- 2\,E_{\underline{\beta},\sigma^2}\{\frac{\phi(F)}{F}\tilde{\underline{\beta}}_2^T\underline{G}(\tilde{\underline{\beta}}_2 - \underline{\beta}_2)\} + E_{\underline{\beta},\sigma^2}\{(\frac{\phi^2(F)}{F^2})\tilde{\underline{\beta}}_2^T\underline{G}\tilde{\underline{\beta}}_2\} \end{aligned} \tag{2.18}$$

Write $\underline{X}\,\tilde{\underline{\beta}} = \underline{X}_1\,\hat{\underline{\beta}}_1 + (\underline{X}_2 - \underline{X}_1\underline{C}_{11}^{-1}\underline{C}_{12})\tilde{\underline{\beta}}_2$. It follows that under the loss (2.5),

$$R(\underline{X}\,\underline{\beta}, \underline{X}\,\tilde{\underline{\beta}}) = tr\,\{(\underline{X}_1^T\underline{Q}\underline{X}_1)V(\hat{\underline{\beta}}_1)\} + E_{\underline{\beta},\sigma^2}\{(\tilde{\underline{\beta}}_2 - \underline{\beta}_2)^T\underline{G}(\tilde{\underline{\beta}}_2 - \underline{\beta}_2)\} \tag{2.19}$$

A generalization of Stein's identity (Berger and Haff, 1983) gives

$$
\begin{aligned}
& E_{\underline{\beta},\sigma^2}\{\frac{\phi(F)}{F}\,\tilde{\underline{\beta}}_2^T\,\underline{G}(\tilde{\underline{\beta}}_2-\underline{\beta}_2)\} \\
= \; & E_{\underline{\beta},\sigma^2}\{tr\,\{[\,\frac{\partial}{\partial\,\tilde{\underline{\beta}}_2}(\frac{\phi(F)}{F}(\underline{G}\tilde{\underline{\beta}}_2))^T\,]\,V(\tilde{\underline{\beta}}_2)\}\} \\
= \; & E_{\underline{\beta},\sigma^2}\{\sigma^2 tr\,\{\frac{\phi(F)}{F}\underline{G}\underline{C}_{22.1}^{-1}\} \\
+ \; & 2\sigma^2(\frac{\phi'(F)}{F}-\frac{\phi(F)}{F^2})(\tilde{\underline{\beta}}_2^T\,\underline{G}\tilde{\underline{\beta}}_2)/((n-p+2)^{-1}SSE)\} \\
= \; & \sigma^2 E_{\underline{\beta},\sigma^2}\{\frac{\phi(F)}{F}tr\,(\underline{G}\,\underline{C}_{22.1}^{-1})+2\,(\phi'(F)-\frac{\phi(F)}{F})\,\frac{\tilde{\underline{\beta}}_2^T\,\underline{G}\tilde{\underline{\beta}}_2}{\tilde{\underline{\beta}}_2^T\,\underline{C}_{22.1}\tilde{\underline{\beta}}_2}\}. \qquad (2.20)
\end{aligned}
$$

Combining (2.18) (2.19) and (2.20), one gets (2.14).

The following corollary to Theorem 2.2.1 provides sufficient conditions under which $\underline{e}_\phi$ dominates $\underline{X}\,\tilde{\underline{\beta}}$. Let $ch_1(\underline{M})$ denote the largest eigenvalue of a matrix $\underline{M}$.

Corollary 2.2.1. Suppose (i) $tr\,(\underline{G}\,\underline{C}_{22.1}^{-1}) > 2ch_1(\underline{G}\,\underline{C}_{22.1}^{-1})$; (ii) $\phi(F)$ is nondecreasing in F and $0<\phi(F)<2\,\{(tr\,(\underline{G}\,\underline{C}_{22.1}^{-1})/ch_1(\underline{G}\,\underline{C}_{22.1}^{-1}))-2\}$. Then $R(\underline{X}\,\underline{\beta},\underline{e}_\phi) < R(\underline{X}\,\underline{\beta},\underline{X}\,\tilde{\underline{\beta}})$.

Proof of Corollary 2.2.1. Following Eqs. (3.8) and (3.9) of Ghosh, Saleh, and Sen (1987), one gets

$$
(\phi^2(F)/F^2)\tilde{\underline{\beta}}_2^T\,\underline{G}\,\tilde{\underline{\beta}}_2 \le ch_1(\underline{G}\,\underline{C}_{22.1}^{-1})h^2(F)F(\frac{n-p}{n-p+2}S_e^2), \qquad (2.21)
$$

where $h(F)=\phi(F)/F$. Now, applying (2.18) of Efron and Morris (1976), one gets

$$
E\{h^2(F)F\frac{n-p}{n-p+2}S_e^2\} = \sigma^2 E\{\frac{\phi^2(F)}{F}-\frac{4}{n-p+2}\phi(F)\phi'(F)\} \qquad (2.22)
$$

From (2.21) and (2.22), one can conclude that

$$
E(\phi^2(F)/F^2)\tilde{\underline{\beta}}_2^T\,\underline{G}\,\tilde{\underline{\beta}}_2 \le ch_1(\underline{G}\,\underline{C}_{22.1}^{-1})\sigma^2 E\{\frac{\phi^2(F)}{F}-\frac{4}{n-p+2}\phi(F)\phi'(F)\}. \qquad (2.23)
$$

Using first part of condition (ii), Eq.(2.14), $\tilde{\underline{\beta}}_2^T\,\underline{G}\,\tilde{\underline{\beta}}_2 \le ch_1(\underline{G}\,\underline{C}_{22.1}^{-1})\{F\frac{n-p}{n-p+2}S_e^2\}$, and Eq.(2.23), it follows that

$$
\begin{aligned}
& R(\underline{X}\,\underline{\beta}, \underline{e}_\phi) - R(\underline{X}\,\underline{\beta}, \underline{X}\,\tilde{\underline{\beta}}) \\
= \; & E_{\underline{\beta},\sigma^2}\{-2\sigma^2 \frac{\phi(F)}{F} tr\,(\underline{G}\underline{C}_{22.1}^{-1}) \\
- \; & 4\sigma^2\{\frac{\phi'(F)}{F} - \frac{\phi(F)}{F^2}\}(\tilde{\underline{\beta}}_2^T\,\underline{G}\,\tilde{\underline{\beta}}_2/(\frac{n-p}{n-p+2}S_e^2)) + \frac{\phi^2(F)}{F^2}\tilde{\underline{\beta}}_2^T\,\underline{G}\,\tilde{\underline{\beta}}_2\} \\
\leq \; & \sigma^2 E_{\underline{\beta},\sigma^2}\{-2\sigma^2(\frac{\phi(F)}{F})tr\,(\underline{G}\underline{C}_{22.1}^{-1}) + 4\frac{\phi(F)}{F}ch_1(\underline{G}\underline{C}_{22.1}^{-1}) + \frac{\phi^2(F)}{F}ch_1(\underline{G}\underline{C}_{22.1}^{-1})\} \\
= \; & \sigma^2 E_{\underline{\beta},\sigma^2}\{\frac{\phi(F)}{F}[-2tr\,(\underline{G}\underline{C}_{22.1}^{-1}) + 4ch_1(\underline{G}\underline{C}_{22.1}^{-1}) + \phi(F)ch_1(\underline{G}\underline{C}_{22.1}^{-1})]\}. \qquad (2.24)
\end{aligned}
$$

Now using condition (i), and the second part of condition (ii), the proof is completed.

Consider the the special case $\underline{Q} = \underline{V}^{-1}$. Then $\underline{G} = \underline{C}_{22.1}$. The following corollary is now immediate from Corollary 2.2.1.

Corollary 2.2.2. Suppose $\underline{Q} = \underline{V}^{-1}$ and $p_2 \geq 3$; $\phi(F)$ is nondecreasing in F and $0 < \phi(F) < 2(p_2 - 2)$. Then,

$$R(\underline{X}\,\underline{\beta}, \underline{e}_\phi) < R(\underline{X}\,\underline{\beta}, \underline{X}\,\tilde{\underline{\beta}}).$$

Corollary 2.2.1 and 2.2.2 are useful in many ways. Note the under the loss $\sigma^{-2}L(\underline{X}\,\underline{\beta}, \underline{a})$, changing the risk only by a constant multiplier, $\underline{X}\,\tilde{\underline{\beta}}$ is a constant risk minimax estimator of $\underline{X}\,\underline{\beta}$. Hence, any estimator dominating $\underline{X}\,\tilde{\underline{\beta}}$ is also a minimax estimator of $\underline{X}\,\underline{\beta}$. Thus Corollary 2.2.1 and 2.2.2 provide useful minimax estimators of $\underline{X}\,\underline{\beta}$. Since $\underline{e}_{EB}$ is a special case of $\underline{e}_\phi$ with $\phi(F) = p-2$, it follows from Corollary 2.2.2 that $\underline{e}_{EB}$ dominates $\underline{X}\,\tilde{\underline{\beta}}$ and is a minimax estimator if $\underline{Q} = \underline{V}^{-1}$ and $p_2 \geq 3$. In special case $\underline{V} = \underline{I}_n$, $\underline{e}_{EB}$ dominates $\underline{X}\,\tilde{\underline{\beta}}$ under squared error loss.

It is also clear from (2.14) that when $\underline{Q} = \underline{V}^{-1}$ and $\phi(F) = c$, a positive constant, then $R(\underline{X}\,\underline{\beta}, \underline{e}_\phi)$ is minimized when $c = p_2 - 2$. Thus, for $\underline{Q} = \underline{V}^{-1}$, $\underline{e}_{EB}$ is optimal within the class of estimators of the form

$$\underline{X}_1\{\hat{\underline{\beta}}_1 + (1 - cF^{-1})(\tilde{\underline{\beta}}_1 - \hat{\underline{\beta}}_1)\} + (1 - cF^{-1})\underline{X}_2\,\tilde{\underline{\beta}}_2.$$

It will be shown later the Bayesian optimality of $\underline{e}_{EB}$ holds within the same class of estimators irrespective of any p.d. matrix $\underline{Q}$ in the loss.

It is not necessarily true though that $\underline{\varepsilon}_{EB}$ always dominates $\underline{X}_1\hat{\underline{\beta}}_1$ even when $\underline{Q} = \underline{V}^{-1}$. Under the loss (2.5) with $\underline{Q} = \underline{V}^{-1}$ and from (2.9), writing $(\underline{X}_1\hat{\underline{\beta}}_1 - \underline{X}\underline{\beta})$ as $(\underline{X}_1(\hat{\underline{\beta}}_1 - E_{\underline{\beta},\sigma^2}(\hat{\underline{\beta}}_1)) - (\underline{X}_2 - \underline{X}_1\underline{C}_{11}^{-1}\underline{C}_{12})\underline{\beta}_2)$, it is easy to check that for estimating $\underline{X}\underline{\beta}$, $\underline{X}_1\hat{\underline{\beta}}_1$ has risk

$$R(\underline{X}\underline{\beta}, \underline{X}_1\hat{\underline{\beta}}_1) = p_1\sigma^2 + \underline{\beta}_2^T\underline{C}_{22.1}\underline{\beta}_2. \tag{2.25}$$

Also, it is easy to see from (2.14) that with $\underline{Q} = \underline{V}^{-1}$ and $\phi(F) = p_2 - 2$,

$$R(\underline{X}\underline{\beta}, \underline{\varepsilon}_{EB}) = p\sigma^2 - 2(p_2-2)^2\sigma^2 E_{\underline{\beta},\sigma^2}(F^{-1}) + (p_2-2)^2 E_{\underline{\beta},\sigma^2}(F^{-2}SSH). \tag{2.26}$$

Since $F = (n-p+2)SSH/SSE$, where for fixed $\underline{\beta}$ and σ^2, SSH and SSE are independently distributed with $SSH \sim \sigma^2\chi^2_{p_2}(\lambda)$, $\lambda = \frac{1}{2}\underline{\beta}_2^T\underline{C}_{22.1}\underline{\beta}_2/\sigma^2$, and $SSE \sim \sigma^2\chi^2_{n-p}$, it follows from (2.26) that

$$R(\underline{X}\underline{\beta}, \underline{\varepsilon}_{EB}) = p\sigma^2 - \sigma^2(p_2-2)^2\{\frac{n-p}{n-p+2}\}E(p_2-2+2K)^{-1}, \tag{2.27}$$

where $K \sim Poisson(\lambda)$. Thus from (2.25) and (2.27), as $\lambda \to 0$, $R(\underline{X}\underline{\beta}, \underline{X}_1\hat{\underline{\beta}}_1) \to p_1\sigma^2$ which $R(\underline{X}\underline{\beta}, \underline{\varepsilon}_{EB}) \to \sigma^2(p-(p_2-2)(n-p)/(n-p+2)) > p_1\sigma^2$ since $p_1+p_2 = p$. On the other hand, as $\lambda \to \infty$, $R(\underline{X}\underline{\beta}, \underline{X}_1\hat{\underline{\beta}}_1) \to \infty$, but $R(\underline{X}\underline{\beta}, \underline{\varepsilon}_{EB}) \to p\sigma^2$. Hence, neither $\underline{X}_1\hat{\underline{\beta}}_1$ nor $\underline{\varepsilon}_{EB}$ dominates the other. The above phenomenon is natural to be expected. Small λ signifies $\underline{\beta}_2$ close to zero, in which case clearly $\underline{X}_1\hat{\underline{\beta}}_1$ is the desired estimator. On the other hand, large λ signifies substantial departure from $H_0 : \underline{\beta}_2 = \underline{0}$ in which case the reduced model is expected to perform poorly, but the EB estimator turns out to be fairly robust. Note that $\underline{X}_1\hat{\underline{\beta}}_1$ is not a minimax estimator of $\underline{X}\underline{\beta}$, and since $R(\underline{X}\underline{\beta}, \underline{X}_1\hat{\underline{\beta}}_1) \to \infty$ as $\lambda \to \infty$, from the robustness criterion, $\underline{\varepsilon}_{EB}$ is clearly the winner over $\underline{X}\tilde{\underline{\beta}}$ or $\underline{X}_1\hat{\underline{\beta}}_1$ when $\underline{Q} = \underline{V}^{-1}$.

The EB estimator also dominates the estimator of $\underline{X}\underline{\beta}$, which was proposed by Sclove (1968), in the frequentist sense. The estimator was developed under the situation that at least p_1 regression coefficients should be included in the model and are willing to estimate $p_2 \geq 3$ more coefficients $\beta_{p_1+1}, \cdots, \beta_p$. For example, in

polynomial regression one may believe that the degree of the polynomial $E(Y_i)$ is at least p_1. The corresponding estimator is of the form:

$$\underline{e}_{Sclove} = \underline{X}_1 \tilde{\underline{\beta}}_1 + (1 - \frac{c}{F})\underline{X}_2 \tilde{\underline{\beta}}_2, \tag{2.28}$$

where c is any number between 0 and $2(p_2 - 2)$ and F as well as $\tilde{\underline{\beta}}$ are defined as before.

Theorem 2.2.2. Suppose $\underline{Q} = \underline{V}^{-1}$ and $p_2 \geq 3$; $\phi(F) = c$, a positive constant, and $0 \leq c \leq 2(p_2 - 2)$. Then,

$$R(\underline{X}\underline{\beta}, \underline{e}_\phi) < R(\underline{X}\underline{\beta}, \underline{e}_{Sclove}). \tag{2.29}$$

Proof of Theorem 2.2.2. Writing

$$\underline{e}_{Sclove} - \underline{X}\underline{\beta} = \underline{X}\tilde{\underline{\beta}} - \underline{X}\underline{\beta} - \frac{c}{F}\underline{X}_2 \tilde{\underline{\beta}}_2,$$

then

$$R(\underline{X}\underline{\beta}, \underline{e}_{Sclove}) = p\sigma^2 - 2\, E_{\underline{\beta},\sigma^2}\{\frac{c}{F}(\underline{X}\tilde{\underline{\beta}} - \underline{X}\underline{\beta})^T \underline{V}^{-1}\underline{X}_2 \tilde{\underline{\beta}}_2 + \frac{c^2}{F^2}\tilde{\underline{\beta}}_2^T \underline{C}_{22}\tilde{\underline{\beta}}_2\}. \tag{2.30}$$

It can write $\underline{X}\tilde{\underline{\beta}} - \underline{X}\underline{\beta}$ in the cross-product term as $\underline{X}_1(\hat{\underline{\beta}}_1 - E(\hat{\underline{\beta}}_1)) + (\underline{X}_2 - \underline{X}_1\underline{C}_{11}^{-1}\underline{C}_{12})(\tilde{\underline{\beta}}_2 - \underline{\beta}_2)$, use the independence of $\hat{\underline{\beta}}_1$ and $\tilde{\underline{\beta}}_2$, and get

$$E_{\underline{\beta},\sigma^2}\{\frac{c}{F}(\underline{X}\tilde{\underline{\beta}} - \underline{X}\underline{\beta})^T \underline{V}^{-1}\underline{X}_2 \tilde{\underline{\beta}}_2\} = E_{\underline{\beta},\sigma^2}\{\frac{c}{F}\tilde{\underline{\beta}}_2^T \underline{C}_{22.1}(\tilde{\underline{\beta}}_2 - \underline{\beta})\}.$$

A generalization of Stein's identity (Berger and Haff, 1983) gives

$$\begin{aligned}
& E_{\underline{\beta},\sigma^2}\{\frac{c}{F}\tilde{\underline{\beta}}_2^T \underline{C}_{22.1}(\tilde{\underline{\beta}}_2 - \underline{\beta})\} \\
= \; & E_{\underline{\beta},\sigma^2}\{tr[(\frac{\partial}{\partial \tilde{\underline{\beta}}_2}(\frac{c}{F}\underline{C}_{22.1}\tilde{\underline{\beta}}_2))V(\tilde{\underline{\beta}}_2)]\} \\
= \; & E_{\underline{\beta},\sigma^2}\{\sigma^2[(\frac{c}{F})p_2 + (-\frac{2c}{F^2})SSH/(n-p+2)^{-1}SSE]\} \\
= \; & \sigma^4(\frac{n-p}{n-p+2})c(p_2-2)E_{\underline{\beta},\sigma^2}\{\frac{1}{SSH}\}.
\end{aligned} \tag{2.31}$$

From (2.30), one gets

$$
\begin{aligned}
& R(\underline{X}\,\underline{\beta}, \underline{e}_{Sclove}) \\
= \; & p\sigma^2 - \sigma^4\left(\frac{n-p}{n-p+2}\right)2c(p_2-2)E_{\underline{\beta},\sigma^2}\left\{\frac{1}{SSH}\right\} \\
+ \; & \sigma^4\left(\frac{n-p}{n-p+2}\right)c^2 E_{\underline{\beta},\sigma^2}\left\{\frac{\tilde{\underline{\beta}}_2^T \underline{C}_{22}\tilde{\underline{\beta}}_2}{SSH^2}\right\} \\
= \; & p\sigma^2 + \sigma^4\left(\frac{n-p}{n-p+2}\right)(-2c(p_2-2)+c^2)E_{\underline{\beta},\sigma^2}\left\{\frac{1}{SSH}\right\} \\
+ \; & \sigma^4\left(\frac{n-p}{n-p+2}\right)c^2 E_{\underline{\beta},\sigma^2}\left\{\frac{\tilde{\underline{\beta}}_2^T \underline{C}_{21}\underline{C}_{11}^{-1}\underline{C}_{12}\tilde{\underline{\beta}}_2}{SSH^2}\right\}. \qquad (2.32)
\end{aligned}
$$

Using (2.14) with $\underline{Q} = \underline{V}^{-1}$, $p_2 \geq 3$ and $\phi(F) = c$, it can obtain $R(\underline{X}\,\underline{\beta}, \underline{e}_\phi) = p\sigma^2 + \sigma^4\left(\frac{n-p}{n-p+2}\right)(-2c(p_2-2)+c^2)\; E_{\underline{\beta},\sigma^2}\left\{\frac{1}{SSH}\right\}$. Thus, the proof is completed.

However, the criticism levelled earlier in this section against $\underline{e}_{EB}$ still remains. This estimator is obtained by estimating B, the Bayes shrinking factor, by a random variable which can assume values bigger than 1 with positive (sometimes, substantially bigger than zero) probability. As mentioned earlier, this deficiency is rectified by $\underline{e}_{EB}^+$. For every estimator $\underline{e}_\phi$ defined in (2.16), there is a corresponding estimator $\underline{e}_\phi^+$ given by

$$
\underline{e}_\phi^+(\underline{Y}) = \underline{X}_1\{\hat{\underline{\beta}}_1 + (1 - \frac{\phi(F)}{F})^+(\tilde{\underline{\beta}}_1 - \hat{\underline{\beta}}_1)\} + (1 - \frac{\phi(F)}{F})^+ \underline{X}_2\,\tilde{\underline{\beta}}_2. \qquad (2.33)
$$

It is natural to anticipate that $\underline{e}_\phi^+$ has smaller risk than $\underline{e}_\phi^+$. The following theorem provides a rigorous justification of this intuition.

Theorem 2.2.3. Under the loss (2.5),

$$
R(\underline{X}\,\underline{\beta}, \underline{e}_\phi^+) < R(\underline{X}\,\underline{\beta}, \underline{e}_\phi). \qquad (2.34)
$$

Proof of Theorem 2.2.3. Note that since $(1 - \frac{\phi(F)}{F})^+$ is differentiable a.e., using $\underline{e}_\phi^+(\underline{Y}) = \underline{X}_1\,\hat{\underline{\beta}}_1 + (1 - \frac{\phi(F)}{F})^+(\underline{X}_2 - \underline{X}_1\,\underline{C}_{11}^{-1}\underline{C}_{12})\tilde{\underline{\beta}}_2$, and $E_{\underline{\beta},\sigma^2}(\hat{\underline{\beta}}_1) = \underline{\beta}_1 + \underline{C}_{11}^{-1}\underline{C}_{12}\underline{\beta}_2$, and, by the independence of $\hat{\underline{\beta}}_1$ and $\tilde{\underline{\beta}}_2$, one gets

$$
\begin{aligned}
& R(\underline{X}\,\underline{\beta}, \underline{e}_\phi^+) \\
= \; & E_{\underline{\beta},\sigma^2}\{(\hat{\underline{\beta}}_1 - E_{\underline{\beta},\sigma^2}(\hat{\underline{\beta}}_1))^T(\underline{X}_1^T\underline{Q}\,\underline{X}_1)(\hat{\underline{\beta}}_1 - E_{\underline{\beta},\sigma^2}(\hat{\underline{\beta}}_1)) \\
+ \; & ((1-\frac{\phi(F)}{F})^+)^2\tilde{\underline{\beta}}_2^T\underline{G}\,\tilde{\underline{\beta}}_2 - 2\,(1-\frac{\phi(F)}{F})^+\tilde{\underline{\beta}}_2^T\underline{G}\,\underline{\beta}_2\} + \underline{\beta}_2^T\underline{G}\,\underline{\beta}_2.
\end{aligned}
$$

$R(\underline{X}\,\underline{\beta}, \underline{e}_\phi)$ can be expressed to the similar form. Then,

$$
\begin{aligned}
& R(\underline{X}\,\underline{\beta}, \underline{e}_\phi) - R(\underline{X}\,\underline{\beta}, \underline{e}_\phi^+) \\
= \; & E_{\underline{\beta},\sigma^2}\{[\,(1-\frac{\phi(F)}{F})^2 - ((1-\frac{\phi(F)}{F})^+)^2\,]\tilde{\underline{\beta}}_2^T\underline{G}\,\tilde{\underline{\beta}}_2 \\
- \; & 2\,[\,(1-\frac{\phi(F)}{F}) - (1-\frac{\phi(F)}{F})^+\,]\tilde{\underline{\beta}}_2^T\underline{G}\,\underline{\beta}_2\} \\
\geq \; & -2\,E_{\underline{\beta},\sigma^2}\{(1-\frac{\phi(F)}{F})I_{(\frac{\phi(F)}{F}>1)}\tilde{\underline{\beta}}_2^T\underline{G}\,\underline{\beta}_2\}. \qquad (2.35)
\end{aligned}
$$

In view of (2.35), for proving the result, it suffices to show that

$$
E_{\underline{\beta},\sigma^2}\{\tilde{\underline{\beta}}_2^T\underline{G}\,\underline{\beta}_2 | SSH = h, SSE = e,\ \exists\ \frac{\phi(F)}{F} > 1\} > 0, \qquad (2.36)
$$

for all $h > 0$ and $e > 0$. However ($\tilde{\underline{\beta}}_2$, SSH) is distributed independently of SSE, and F is a multiple of SSH/SSE. Hence, it suffices to show that

$$
E_{\underline{\beta},\sigma^2}\{\tilde{\underline{\beta}}_2^T\underline{G}\,\underline{\beta}_2 | SSH = h\} > 0\ \forall\ h\ > 0. \qquad (2.37)
$$

Since $\underline{C}_{22.1}$ is p.d., there exists a nonsingular $\underline{K}$ such that $\underline{C}_{22.1} = \underline{K}^T\underline{K}$. Write $\underline{Z} = \underline{K}\,\tilde{\underline{\beta}}_2$. Then $SSH = \underline{Z}^T\,\underline{Z}$ and $\underline{Z} \sim N(\underline{K}\,\underline{\beta}_2, \sigma^2\underline{I}_{p_2})$. One can write $\tilde{\underline{\beta}}_2^T\underline{G}\underline{\beta}_2 = (\underline{K}\,\tilde{\underline{\beta}}_2)^T\underline{K}^{-1}\underline{G}\,\underline{\beta}_2 = \underline{Z}^T\underline{\phi}$ (say), $\underline{\phi} = \underline{K}^{-1}\underline{G}\,\underline{\beta}_2$. Now using an orthogonal transformation $\underline{U} = \underline{P}\,\underline{Z}$, where the first row of $\underline{P}$, is $\underline{\phi}^T/\parallel\underline{\phi}\parallel$, one get from (2.37),

$$
E_{\underline{\beta},\sigma^2}\{\tilde{\underline{\beta}}_2^T\underline{G}\,\underline{\beta}_2 | SSH = h\} = \parallel\underline{\phi}\parallel E_{\underline{\beta},\sigma^2}\{U_1 |\ \parallel\underline{U}\parallel^2 = h\}, \qquad (2.38)
$$

where U_1 is the first element of $\underline{U}$ and $\underline{U} \sim N(\underline{\theta}, \sigma^2\underline{I}_{p_2})$, $\underline{\theta} = \underline{P}\,\underline{K}\,\underline{\beta}_2$. From Eq.(2.38), arguing as in Theorem 6.2 (pp 302-303) of Lehmann (1983), one gets $E_{\underline{\beta},\sigma^2}\{U_1 |\ \parallel\underline{U}\parallel^2 = h\} > 0$ and completes the proof of Theorem 2.2.3.

Thus $\underline{e}_{EB}^+$ has smaller risk than $\underline{e}_{EB}$. In Section 2.3, we shall find certain HB estimators which are also of the form $\underline{e}_\phi^+$. Although, these positive part estimators

being non-smooth, are inadmissible under quadratic loss, no explicit estimators dominating these positive-part estimators are available as yet. In any case, one does not expect any significant savings in the risk over the positive part estimators.

The discussion of the frequentist risk is concluded by an introduction of the PTE's. A general PTE is given by

$$\underline{\delta}_{PTE} = \underline{X}_1 [\, \underline{\hat{\beta}}_1 + g(F)(\underline{\tilde{\beta}}_1 - \underline{\hat{\beta}}_1) \,] + \underline{X}_2\, g(F) \underline{\tilde{\beta}}_2, \tag{2.39}$$

where $g(F) = I_{[F>d]}$, d being a positive constant depending on the chosen level of significance, and I the usual indicator function. We shall see, however, that the modified EB estimator

$$\underline{\delta}^{\phi_0}_{MEB}(\underline{Y}) = \underline{X}_1 [\underline{\hat{\beta}}_1 + (1 - \frac{\phi_0(F)}{F}) g(F)(\underline{\tilde{\beta}}_1 - \underline{\hat{\beta}}_1) \,] + \underline{X}_2\, (1 - \frac{\phi_0(F)}{F}) g(F) \underline{\tilde{\beta}}_2, \tag{2.40}$$

dominates $\underline{\delta}_{PTE}$ under certain conditions. Recall the definition of $\underline{G}$ in Theorem 2.2.1. The following theorem is proved.

Theorem 2.2.4. Consider the loss (2.5). Then, if conditions (i) and (ii) of Corollary 2.2.1 hold with $\phi_0(F)$ replacing $\phi(F)$, then,

$$R(\underline{X}\,\underline{\beta}, \underline{\delta}^{\phi_0}_{MEB}) < R(\underline{X}\,\underline{\beta}, \underline{\delta}_{PTE}). \tag{2.41}$$

If $\underline{Q} = \underline{V}^{-1}$ and $p_2 \geq 3$, then (2.41) holds under the conditions of Corollary 2.2.2 with $\phi_0(F)$ replacing $\phi(F)$.

Proof of Theorem 2.2.4. One writes $\underline{\delta}_{PTE} = \underline{e}_{\phi_1}$, where $\phi_1(F) = F(1 - g(F))$ and $\underline{\delta}_{MEB} = \underline{e}_{\phi_2}$ where $\phi_2(F) = F[\, 1 - (1 - \frac{\phi_0(F)}{F}) g(F) \,] = \phi_1(F) + \phi_0(F) g(F)$. Since $g(F)$ is differenable except at $F = d$ and $g'(F) = 0$ a.e., $\phi_1'(F) = \phi_2'(F) - \phi_0'(F) g(F)$ a.e.. Also, $\phi_1(F) - \phi_2(F) = -\phi_0(F) g(F)$ and $\phi_1^2(F) - \phi_2^2(F) = -2\phi_0(F)\phi_1(F) g(F) - \phi_0^2(F) g^2(F) = -\phi_0^2(F) g(F)$ (since $g(F)(1 - g(F)) = 0$ and $g^2(F) = g(F)$). Now,

using (2.14), it can be shown that

$$
\begin{aligned}
& R(\underline{X}\,\underline{\beta}, \underline{\delta}_{PTE}) - R(\underline{X}\,\underline{\beta}, \underline{\delta}_{MEB}) \\
= \; & R(\underline{X}\,\underline{\beta}, \underline{e}_{\phi_1}) - R(\underline{X}\,\underline{\beta}, \underline{e}_{\phi_2}) \\
= \; & -4\sigma^2 E_{\underline{\beta},\sigma^2}\{-\phi_0'(F)g(F)\frac{\tilde{\underline{\beta}}_2^T \underline{G}\, \tilde{\underline{\beta}}_2}{SSH}\} \\
& -2\sigma^2 E_{\underline{\beta},\sigma^2}\{\frac{-\phi_0(F)g(F)}{F}[\,tr\,(\underline{G}\,\underline{C}_{22.1}^{-1}) - 2\frac{\tilde{\underline{\beta}}_2^T \underline{G}\, \tilde{\underline{\beta}}_2}{SSH}]\} \\
& +E_{\underline{\beta},\sigma^2}\{-\frac{\phi_0^2(F)}{F^2}g(F)\tilde{\underline{\beta}}_2^T \underline{G}\, \tilde{\underline{\beta}}_2\}. \qquad (2.42)
\end{aligned}
$$

Since $\phi_0'(F) \geq 0$ a.e., it follows from (2.42) that

$$
\begin{aligned}
& R(\underline{X}\,\underline{\beta}, \underline{\delta}_{PTE}) - R(\underline{X}\,\underline{\beta}, \underline{\delta}_{MEB}) \\
\geq \; & 2\sigma^2 E_{\underline{\beta},\sigma^2}\{\frac{\phi_0(F)g(F)}{F}[\,tr\,(\underline{G}\,\underline{C}_{22.1}^{-1}) - 2\frac{\tilde{\underline{\beta}}_2^T \underline{G}\, \tilde{\underline{\beta}}_2}{SSH}]\} \\
- \; & E_{\underline{\beta},\sigma^2}\{\frac{\phi_0^2(F)}{F^2}g(F)\tilde{\underline{\beta}}_2^T \underline{G}\, \tilde{\underline{\beta}}_2\}. \qquad (2.43)
\end{aligned}
$$

Arguing as in Corollary 2.2.1, one gets (2.41) from (2.43).

In the particular case $\underline{Q} = \underline{V}^{-1}$ so that $\underline{G} = \underline{C}_{22.1}$, it follows from (2.43) that

$$
\begin{aligned}
& R(\underline{X}\,\underline{\beta}, \underline{\delta}_{PTE}) - R(\underline{X}\,\underline{\beta}, \underline{\delta}_{MEB}) \\
\geq \; & 2(p_2 - 2)\sigma^2 E_{\underline{\beta},\sigma^2}\{\frac{\phi_0(F)g(F)}{F}\} - E_{\underline{\beta},\sigma^2}\{\frac{\phi_0^2(F)}{F}g(F)\frac{SSE}{n-p+2}\}. \qquad (2.44)
\end{aligned}
$$

Now define $h(F) = \phi_0(F)/F$ and apply (2.18) of Efron and Morris (1976) to get

$$
\begin{aligned}
& E_{\underline{\beta},\sigma^2}\{h^2(F)Fg(F)\frac{SSE}{n-p+2}\} \\
= \; & \sigma^2 E_{\underline{\beta},\sigma^2}\{\frac{n-p}{n-p+2}h^2(F)Fg(F) \\
+ \; & \frac{2}{n-p+2}\cdot\frac{n-p}{n-p+2}S_e^2(2h(F)h'(F)Fg(F) + h^2(F)g(F))(-\frac{F}{\frac{n-p}{n-p+2}S_e^2})\} \\
= \; & \sigma^2 E_{\underline{\beta},\sigma^2}\{\frac{n-p-2}{n-p+2}h^2(F)Fg(F) - \frac{4}{n-p+2}F^2 g(F)h(F)h'(F)\} \\
= \; & \sigma^2 E_{\underline{\beta},\sigma^2}\{\frac{n-p-2}{n-p+2}\cdot\frac{\phi_0^2(F)g(F)}{F} - \frac{4}{n-p+2}Fg(F)\phi_0(F)(\frac{\phi_0'(F)}{F} - \frac{\phi_0(F)}{F^2}\} \\
= \; & \sigma^2 E_{\underline{\beta},\sigma^2}\{\frac{\phi_0^2(F)g(F)}{F} - \frac{4}{n-p+2}\phi_0(F)g(F)\phi_0'(F)\} \\
\leq \; & \sigma^2 E_{\underline{\beta},\sigma^2}\{\frac{\phi_0^2(F)g(F)}{F}\}. \qquad (2.45)
\end{aligned}
$$

Combining Eqs.(2.44) and (2.45), the second part of Theorem 2.2.4 is therefore proved.

Note that $\underline{\delta}^{\phi_0}_{MEB}$ estimates $\underline{X}\underline{\beta}$ by $\underline{X}_1\hat{\underline{\beta}}_1$ if $g(F) = 0$, i.e., one accepts the null hypothesis at a desired level of significance. However, in the event the null hypothesis is rejected, one estimats $\underline{X}\underline{\beta}$ by a shrinkage estimator as described in the earlier part of this section instead of $\underline{X}\tilde{\underline{\beta}}$.

The above theorem provides a class of estimators improving on the PTE. However, in view of Theorem 2.2.3, $\underline{\delta}^{\phi_0}_{MEB}$ can be further improved by a positive part modified EB estimator $(\underline{\delta}^{\phi_0}_{MEB})^+$ given by

$$
\begin{aligned}
&(\underline{\delta}^{\phi_0}_{MEB})^+(\underline{Y}) \\
&= \underline{X}_1[\hat{\underline{\beta}}_1 + (1 - \frac{\phi_0(F)}{F})^+ g(F)(\tilde{\underline{\beta}}_1 - \hat{\underline{\beta}}_1)] + \underline{X}_2(1 - \frac{\phi_0(F)}{F})^+ g(F)\tilde{\underline{\beta}}_2. \quad (2.46)
\end{aligned}
$$

Now, the Bayes risk optimality of $\underline{e}_{EB}$ is discussed. In view of Theorem 2.2.1, the Bayes risk of $\underline{e}_{EB}$ is smaller than that of $\underline{X}\tilde{\underline{\beta}}$ when $\underline{Q} = \underline{V}^{-1}$ and $p_2 \geq 3$ irrespective of any prior, while $\underline{e}^+_{EB}$ always dominates $\underline{e}_{EB}$ irrespective of any prior. However, the Bayes risk optimality of $\underline{e}_{EB}$ over $\underline{X}\tilde{\underline{\beta}}$ is not restricted to $\underline{Q} = \underline{V}^{-1}$. The following theorem shows that under the prior described at the beginning of this section, the Bayes risk of $\underline{e}_{EB}$ is smaller than that of $\underline{X}\tilde{\underline{\beta}}$ irrespective of any p.d. $\underline{Q}$.

Theorem 2.2.5. Consider the model $\underline{Y}|\underline{\beta} \sim N(\underline{X}\underline{\beta}, \sigma^2\underline{V})$, where $\underline{V}$ is p.d. and $\underline{\beta} \sim N(\underline{\nu}, \tau^2(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1})$ with $\underline{\nu}^T = (\underline{\nu}_1^T, \underline{0}^T)$. Denote the given prior by ξ and let $r(\xi, \underline{e})$ denote the Bayes risk of an estimator $\underline{e}$ of $\underline{X}\underline{\beta}$ under the prior ξ when the loss is given in (2.5). Then,

$$
r(\xi, \underline{e}_{EB}) < r(\xi, \underline{X}\tilde{\underline{\beta}}). \quad (2.47)
$$

Proof of Theorem 2.2.5. From (2.6), $\underline{e}_{EB} - \underline{X}\underline{\beta}$ can be expressed as

$$
\underline{X}(\tilde{\underline{\beta}} - \underline{\beta} - \frac{p_2 - 2}{F}(\tilde{\underline{\beta}} - \begin{pmatrix} \hat{\underline{\beta}}_1 \\ \underline{0} \end{pmatrix})). \quad (2.48)
$$

Hence, one gets

$$
\begin{aligned}
r(\xi, \underline{\varepsilon}_{EB}) &= E\{(\tilde{\underline{\beta}} - \underline{\beta})^T \underline{X}^T \underline{Q}\, \underline{X} (\tilde{\underline{\beta}} - \underline{\beta}) \\
&\quad -2\,\frac{p_2 - 2}{F} (\tilde{\underline{\beta}} - \underline{\beta})^T \underline{X}^T \underline{Q}\, \underline{X} (\tilde{\underline{\beta}} - \begin{pmatrix} \hat{\underline{\beta}}_1 \\ \underline{0} \end{pmatrix}) \\
&\quad +\frac{(p_2 - 2)^2}{F^2} (\tilde{\underline{\beta}} - \begin{pmatrix} \hat{\underline{\beta}}_1 \\ \underline{0} \end{pmatrix})^T \underline{X}^T \underline{Q}\, \underline{X} (\tilde{\underline{\beta}} - \begin{pmatrix} \hat{\underline{\beta}}_1 \\ \underline{0} \end{pmatrix})\}. \qquad (2.49)
\end{aligned}
$$

Since $E_{\underline{\beta},\sigma^2}(\tilde{\underline{\beta}} - \underline{\beta})^T \underline{X}^T \underline{Q}\, \underline{X} (\tilde{\underline{\beta}} - \underline{\beta}) = \sigma^2 tr\,[\,(\underline{X}^T \underline{V}^{-1}\, \underline{X})^{-1} \underline{X}^T \underline{Q}\, \underline{X}\,]$, therefore, under the marginal distribution of $\underline{\beta}$, one gets

$$
E(\tilde{\underline{\beta}} - \underline{\beta})^T \underline{X}^T \underline{Q}\, \underline{X} (\tilde{\underline{\beta}} - \underline{\beta}) = \sigma^2 tr\,[\,(\underline{X}^T \underline{V}^{-1}\, \underline{X})^{-1} \underline{X}^T \underline{Q}\, \underline{X}\,]. \qquad (2.50)
$$

Next note that $\hat{\underline{\beta}}_1$, $\tilde{\underline{\beta}}_2$ and $SSE = (\underline{Y} - \underline{X}\tilde{\underline{\beta}})^T \underline{V}^{-1} (\underline{Y} - \underline{X}\tilde{\underline{\beta}})$ are mutually independent (see Lemma 2 of Ghosh, Saleh and Sen, 1987) and marginally $\hat{\underline{\beta}}_1 \sim N(\underline{\nu}_1, \sigma^2 B^{-1} \underline{C}_{11}^{-1})$ and $\tilde{\underline{\beta}}_2 \sim N(\underline{0}, \sigma^2 B^{-1} \underline{C}_{22.1}^{-1})$, and $SSE \sim \sigma^2 \chi^2_{n-p}$. Recall $B = \sigma^2/(\sigma^2 + \tau^2)$. Then, by conditioning on $\underline{Y}$ and using $E(\underline{\beta}|\underline{Y}) = \tilde{\underline{\beta}} - B(\tilde{\underline{\beta}} - \underline{\nu})$ with $\underline{\nu}^T = (\underline{\nu}_1^T, \underline{0}^T)$ and Eq.(2.9), the cross-term in (2.49) divided by 2 becomes

$$
\begin{aligned}
& E\{\frac{p_2 - 2}{F} (\tilde{\underline{\beta}} - E(\underline{\beta}|\underline{Y}))^T \underline{X}^T \underline{Q}\, \underline{X} \begin{pmatrix} -\underline{C}_{11}^{-1} \underline{C}_{12} \\ \underline{I}_{p_2} \end{pmatrix} \tilde{\underline{\beta}}_2\} \\
= \; & B\, E\{\frac{p_2 - 2}{F} \begin{pmatrix} \hat{\underline{\beta}}_1 - \underline{\nu}_1 - \underline{C}_{11}^{-1} \underline{C}_{12} \tilde{\underline{\beta}}_2 \\ \tilde{\underline{\beta}}_2 \end{pmatrix}^T \underline{X}^T \underline{Q}\, \underline{X} \begin{pmatrix} -\underline{C}_{11}^{-1} \underline{C}_{12} \\ \underline{I}_{p_2} \end{pmatrix} \tilde{\underline{\beta}}_2\} \\
= \; & B\, E\{\frac{p_2 - 2}{F} \tilde{\underline{\beta}}_2^T (-\underline{C}_{21} \underline{C}_{11}^{-1}, \; \underline{I}_{p_2}) \underline{X}^T \underline{Q}\, \underline{X} \begin{pmatrix} -\underline{C}_{11}^{-1} \underline{C}_{12} \\ \underline{I}_{p_2} \end{pmatrix} \tilde{\underline{\beta}}_2\} \\
= \; & B\, E\{\frac{p_2 - 2}{F} \tilde{\underline{\beta}}_2^T \underline{G}\, \tilde{\underline{\beta}}_2\} \\
= \; & \sigma^2 B\, \frac{(p_2 - 2)(n - p)}{n - p + 2} E(\frac{\tilde{\underline{\beta}}_2^T \underline{G}\, \tilde{\underline{\beta}}_2}{\tilde{\underline{\beta}}_2^T \underline{C}_{22.1}\, \tilde{\underline{\beta}}_2}). \qquad (2.51)
\end{aligned}
$$

Now, $\tilde{\underline{\beta}}_2^T \underline{C}_{22.1}\, \tilde{\underline{\beta}}_2$ is a function of the complete sufficient statistic, where $\tilde{\underline{\beta}}_2^T \underline{G}\, \tilde{\underline{\beta}}_2 \,/ \tilde{\underline{\beta}}_2^T\, \underline{C}_{22.1} \tilde{\underline{\beta}}_2$ is ancillary. Hence, using Basu's Theorem,

$$
E(\tilde{\underline{\beta}}_2^T \underline{G}\, \tilde{\underline{\beta}}_2) = E(\tilde{\underline{\beta}}_2^T \underline{C}_{22.1} \tilde{\underline{\beta}}_2) E(\tilde{\underline{\beta}}_2^T \underline{G}\, \tilde{\underline{\beta}}_2 / \tilde{\underline{\beta}}_2^T \underline{C}_{22.1} \tilde{\underline{\beta}}_2).
$$

From the above identity and the fact that $\tilde{\underline{\beta}}_2 \sim N(\underline{0}, \sigma^2 \underline{C}_{22.1}^{-1})$, one gets

$$
E\frac{\tilde{\underline{\beta}}_2^T \underline{G}\, \tilde{\underline{\beta}}_2}{\tilde{\underline{\beta}}_2^T \underline{C}_{22.1}\, \tilde{\underline{\beta}}_2} = \frac{E(\tilde{\underline{\beta}}_2^T \underline{G}\, \tilde{\underline{\beta}}_2)}{E(\tilde{\underline{\beta}}_2^T \underline{C}_{22.1}\, \tilde{\underline{\beta}}_2)} = \frac{tr\,(\underline{G}\, \underline{C}_{22.1}^{-1})}{p_2}. \qquad (2.52)
$$

Using (2.9), the third term of (2.49) can be written as

$$
\begin{aligned}
& E\{\frac{(p_2-2)^2}{F^2}\tilde{\underline{\beta}}_2^T(-\underline{C}_{21}\underline{C}_{11}^{-1},\ \underline{I}_{p_2})\underline{X}^T\underline{Q}\,\underline{X}\begin{pmatrix} -\underline{C}_{11}^{-1}\underline{C}_{12} \\ \underline{I}_{p_2}\end{pmatrix}\tilde{\underline{\beta}}_2\} \\
= \ & E\{\frac{(p_2-2)^2}{F^2}\tilde{\underline{\beta}}_2^T\underline{G}\,\tilde{\underline{\beta}}_2\} \\
= \ & \sigma^4\frac{(p_2-2)^2(n-p)}{(n-p+2)}E\{\frac{\tilde{\underline{\beta}}_2^T\underline{G}\,\tilde{\underline{\beta}}_2}{(\tilde{\underline{\beta}}_2^T\underline{C}_{22.1}\,\tilde{\underline{\beta}}_2)^2}\}. \qquad (2.53)
\end{aligned}
$$

Finally, using the similar argument of Eq.(2.52),

$$
\begin{aligned}
& E\{\frac{\tilde{\underline{\beta}}_2^T\underline{G}\,\tilde{\underline{\beta}}_2}{(\tilde{\underline{\beta}}_2^T\underline{C}_{22.1}\,\tilde{\underline{\beta}}_2)^2}\} \\
= \ & E\{\frac{\tilde{\underline{\beta}}_2^T\underline{G}\,\tilde{\underline{\beta}}_2}{\tilde{\underline{\beta}}_2^T\underline{C}_{22.1}\,\tilde{\underline{\beta}}_2}\}E(\tilde{\underline{\beta}}_2^T\underline{C}_{22.1}\,\tilde{\underline{\beta}}_2)^{-1} \\
= \ & \{\frac{tr\,[\underline{G}\,\underline{C}_{22.1}^{-1}]}{p_2}\}\sigma^{-2}B(p_2-2)^{-1}. \qquad (2.54)
\end{aligned}
$$

Combine (2.50), (2.52) and (2.54) to get,

$$
r(\xi,\underline{e}_{EB}) = r(\xi,\underline{X}\,\tilde{\underline{\beta}}) - \sigma^2 B\frac{n-p}{n-p+2}\frac{p_2-2}{p_2}tr\,[\underline{G}\,\underline{C}_{22.1}^{-1}]. \qquad (2.55)
$$

It completes the proof of (2.47).

The next theorem provides necessary and suffcient conditions under which the Bayes risk of $\underline{e}_{EB}$ is smaller than that of $\underline{X}_1\hat{\underline{\beta}}_1$.

Theorem 2.2.6. Consider the set of Theorem 2.2.5.

$$
r(\xi,\underline{e}_{EB}) < r(\xi,\underline{X}_1\hat{\underline{\beta}}_1), \qquad (2.56)
$$

if and only if $(1-B)^2/B^2 > 1-\{(n-p)/(n-p+2)\}\{(p_2-2)/p_2\}$.

Proof of Theorem 2.2.6. Using the identity of $(\underline{X}_1\hat{\underline{\beta}}_1-\underline{X}\,\underline{\beta}) = \underline{X}_1(\hat{\underline{\beta}}_1 - E_{\underline{\beta},\sigma^2}(\hat{\underline{\beta}}_1)) - (\underline{X}_2 - \underline{X}_1\underline{C}_{11}^{-1}\underline{C}_{12})\underline{\beta}_2$, one gets

$$
R(\underline{X}\,\underline{\beta},\underline{X}_1\hat{\underline{\beta}}_1) = \sigma^2 tr\,((\underline{X}_1^T\underline{Q}\,\underline{X}_1)\underline{C}_{11}^{-1}) + \underline{\beta}_2^T\underline{G}\,\underline{\beta}_2. \qquad (2.57)
$$

Recall $\underline{\beta} \sim N(\underline{\nu}, \tau^2(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1})$ with $\underline{\nu}^T = (\underline{\nu}_1^T, \underline{0}^T)$. Then

$$r(\xi, \underline{X}_1\underline{\hat{\beta}}_{.1}) = \sigma^2 tr\left((\underline{X}_1^T\underline{Q}\,\underline{X}_1)\underline{C}_{11}^{-1}\right) + \tau^2 tr\left(\underline{G}\,\underline{C}_{22.1}^{-1}\right). \tag{2.58}$$

Since $tr\left((\underline{X}_1^T\underline{Q}\,\underline{X}_1)\underline{C}_{11}^{-1}\right) + tr\left(\underline{G}\,\underline{C}_{22.1}^{-1}\right) = tr\left((\underline{X}^T\underline{Q}\underline{X})(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\right)$ after some algebra. Now, from Eqs.(2.58) and (2.55),

$$r(\xi, \underline{X}_1\underline{\hat{\beta}}_{.1}) - r(\xi, \underline{e}_{EB}) = (\tau^2 - \sigma^2 + B\sigma^2\frac{n-p}{n-p+2}\frac{p_2-2}{p_2})tr\left(\underline{G}\,\underline{C}_{22.1}^{-1}\right) > 0, \tag{2.59}$$

occurs as $\tau^2/\sigma^2 > 1 - B(n-p)(p_2-2)/(n-p+2)(p_2)$ i.e. $(1-B)^2/(B^2) > 1 - [(n-p)/(n-p+2)][(p_2-2)/p_2]$.

The significance of the above theorem is very clear. If $\sigma^2 < \tau^2$, then $B < \frac{1}{2}$ in which case $(1-B)^2/B^2 > 1$. Thus $\underline{e}_{EB}$ always dominates $\underline{X}_1\underline{\hat{\beta}}_{.1}$ in the Bayes risk if the sampling variability is smaller than the prior variability. However, if τ^2 is much smaller than σ^2, then the reduced model seems to very appropriate, and $\underline{X}_1\underline{\hat{\beta}}_{.1}$ becomes a clear winner.

As discussed earlier in this section, one can obtain a general class of minimax estimates of $\underline{X}\,\underline{\beta}$ of the form

$$\underline{e}_c(\underline{Y}) = \underline{X}_1\left[\underline{\hat{\beta}}_{.1} + (1 - \frac{c}{F})(\underline{\tilde{\beta}}_{.1} - \underline{\hat{\beta}}_{.1})\right] + \underline{X}_2(1 - \frac{c}{F})\underline{\tilde{\beta}}_{.2}, \tag{2.60}$$

where c is an appropriate constant. Under the same prior ξ, the following theorem shows that the Bayes risk optimality of $\underline{e}_c$ over $\underline{e}_{Sclove}$ when $\underline{Q} = \underline{V}^{-1}$.

Theorem 2.2.7. Under the loss

$$L(\underline{X}\,\underline{\beta}, \underline{a}) = (\underline{a} - \underline{X}\,\underline{\beta})^T\,\underline{V}^{-1}(\underline{a} - \underline{X}\,\underline{\beta}),$$

then

$$r(\xi, \underline{e}_c) < r(\xi, \underline{e}_{Sclove}),$$

where $0 < c < 2(p_2 - 2)$.

Proof of Theorem 2.2.7. From the proof of Theorem 2.2.5, following the procedure of obtaining $r(\xi, \underline{e}_{EB})$ with $\underline{e}_c$ replacing $\underline{e}_{EB}$ and $\underline{V}^{-1}$ replacing $\underline{Q}$ one gets $\underline{G} = \underline{C}_{22.1}$. Hence

$$r(\xi, \underline{e}_c) = p\sigma^2 + (\frac{c^2}{p_2 - 2} - 2c)\sigma^2 B \frac{n-p}{n-p+2}. \tag{2.61}$$

Using the result of $R(\underline{X}\underline{\beta}, \underline{e}_{Sclove})$ in Eq.(2.32), and Basu's theorem, it shows that

$$r(\xi, \underline{e}_{Sclove}) = r(\xi, \underline{e}_c) + c^2\sigma^2 B \frac{n-p}{n-p+2} \frac{tr\,(\underline{C}_{22.1}^{-1}(\underline{C}_{21}\underline{C}_{11}^{-1}\underline{C}_{12}))}{p_2(p_2-2)}. \tag{2.62}$$

The conclusion of the theorem follows.

In fact, if one estimates the prior parameters by some other method than what we have proposed, one can get EB estimators different from $\underline{e}_{EB}$. However, the following theorem shows that $\underline{e}_{EB}$ enjoys an optimality within the class of all estimators of the form $\underline{e}_c(\underline{Y})$, i.e. among those with $\phi(F) = c$ for all F.

Theorem 2.2.8. Consider the same set up as of Theorem 2.2.5. Then,

$$\begin{aligned} r(\xi, \underline{e}_c) - r(\xi, \underline{e}_B) &= \sigma^2 B(tr\{(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\underline{X}^T\underline{Q}\,\underline{X}\}) \\ &+ (\frac{c^2}{p_2-2} - 2c)\sigma^2 B(\frac{n-p}{n-p+2})\frac{tr\,(\underline{G}\,\underline{C}_{22.1}^{-1})}{p_2}. \end{aligned} \tag{2.63}$$

For $p_2 \geq 3$, the above risk is minimized at $c = p_2 - 2$.

Proof of Theorem 2.2.8. With c replacing $p_2 - 2$ in the procedure of obtaining $r(\xi, \underline{e}_{EB})$ in the proof of Theorem 2.2.5, one gets

$$\begin{aligned} r(\xi, \underline{e}_c) &= \sigma^2 tr\,[\,(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}(\underline{X}^T\underline{Q}\,\underline{X})\,] \\ &+ (\frac{c^2}{p_2-2} - 2c)\sigma^2 B(\frac{n-p}{n-p+2})\frac{tr\,(\underline{G}\,\underline{C}_{22.1}^{-1})}{p_2}. \end{aligned} \tag{2.64}$$

And

$$\begin{aligned} r(\xi, \underline{e}_B) &= tr\,\{E(\underline{e}_B - \underline{\beta})(\underline{e}_B - \underline{\beta})^T(\underline{X}^T\underline{Q}\,\underline{X})\} \\ &= tr\,\{E[\,E(\underline{e}_B - \underline{\beta})(\underline{e}_B - \underline{\beta})^T|\underline{Y}\,](\underline{X}^T\underline{Q}\,\underline{X})\} \\ &= tr\,\{V(\underline{\beta}|\underline{Y})(\underline{X}^T\underline{Q}\,\underline{X})\} \\ &= \sigma^2(1-B)tr\,[\,(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}(\underline{X}^T\underline{Q}\,\underline{X})\,]. \end{aligned} \tag{2.65}$$

Combine (2.64) and (2.65) to get (2.63). Thus, the proof of Theorem 2.2.8 is completed.

It is also possible to motivate the proposed $\underline{e}_{EB}$ by using a prior distribution on σ^2 instead of estimating σ^2 from the marginal distribution of $\underline{Y}$. Consider

$$(\underline{Y}|\underline{\beta},\sigma^2) \sim N(\underline{X}\,\underline{\beta},\sigma^2\underline{V}),$$
$$(\underline{\beta}|\sigma^2) \sim N(\begin{pmatrix} \underline{\nu}_1 \\ \underline{0} \end{pmatrix}, \sigma^2 g^{-1}(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1})$$

and σ^2 has some *pdf* $m(\sigma^2)$ (say). Under this prior, the posterior mean of $\underline{X}\,\underline{\beta}$ is still identical to the expression given in (2.6). Note that $B = \sigma^2/(\sigma^2 + \sigma^2 g^{-1}) = g/(1+g)$. It follows from (2.7)

$$(\underline{Y}|\sigma^2) \sim N_n(\underline{X}_1\,\underline{\nu}_1, \sigma^2(\underline{V} + g^{-1}\underline{P}_{\underline{X}})).$$

Accordingly $SSH|\sigma^2 \sim \sigma^2 B^{-1}\chi^2_{p_2}$, while $SSE|\sigma^2 \sim \sigma^2\chi^2_{n-p}$. Thus, it is still sensible to estimate B by $(p_2-2)/F$ for more appropriately by $\min(1,(p_2-2)/F)$. The above EB estimator maintains its frequentist and Bayes risk optimality within the class of estimators $\underline{e}_c$ as given in (2.60) even under this alternative model.

2.3 The Hierarchical Bayes Approach

As discussed in Section 2.2, as an alternative to the classical EB method, one can use a HB method, where one put a prior distribution, often improper, on the unknown hyperparameters, and arrive at the posterior distribution of $\underline{\beta}$ given $\underline{Y} = \underline{y}$. Inference about $\underline{\beta}$ or $\underline{X}\,\underline{\beta}$ is then based on this posterior distribution. For point estimation, usually both these approaches stand more or less on equal footing. However, unlike the EB estimator of $\underline{\beta}$ or $\underline{X}\,\underline{\beta}$ which has no natural confidence or credible set associated with it, the HB method provides, at least with reasonable approximation, a credible set for $\underline{X}\,\underline{\beta}$.

In this section a general posterior distribution of $\underline{\beta}$ given $\underline{Y} = \underline{y}$ is first derived. In the special case of the g-prior given in Section 2.2, the posterior distribution

simplifies considerbly. In that case, for certain improper priors on σ^2 and τ^2, there is considerable similarily between the EB and HB approaches. Also, some of the positive part EB estimators can be identified as HB estimators, where one uses the posterior mode instead of the posterior mean for estimating certain parameters.

In order to provide the first general result of this section, first write $\sigma^2 = r_1^{-1}$. Consider the model under which $(\underline{Y}|\underline{\beta}, r_1) \sim N(\underline{X}\,\underline{\beta}, r_1^{-1}\underline{V})$, where $\underline{V}$ is a known p.d. matrix. Also, let $(\underline{\beta}|\underline{\nu}, r_2) \sim N(\underline{\nu}, r_2^{-1}\underline{D}^{-1})$, where $\underline{\nu}^T = (\underline{\nu}_1^T, \underline{0}^T)$, and $\underline{D}$ is a p.d. matrix. Finally assume that $\underline{\nu}_1$, R_1 and R_2 are marginally independently distributed with $\underline{\nu}_1 \sim$ uniform on R^{p_1}, the p_1—dimensional Euclidean space, $R_1 \sim Gamma(\frac{1}{2}a_1, \frac{1}{2}f_1)$ and $R_2 \sim Gamma(\frac{1}{2}a_2, \frac{1}{2}f_2)$. Z is said to have a $Gamma(\alpha, \beta)$ distribution if Z has *pdf*

$$f(z) = \exp(-\alpha z) z^{\beta-1} \alpha^\beta / \Gamma(\beta) I_{[z>0]},\ \alpha > 0,\ \beta > 0.$$

In order to find the posterior distribution of $\underline{\beta}$, recall that $\underline{C}_{ij} = \underline{X}_i^T \underline{V}^{-1} \underline{X}_j$ $(i, j = 1, 2)$, and $\underline{C}_{22.1} = \underline{C}_{22} - \underline{C}_{21}\,\underline{C}_{22}^{-1}\underline{C}_{12}$. Also, partition $\underline{D}$ as $\underline{D} = \begin{pmatrix} \underline{D}_{11} & \underline{D}_{12} \\ \underline{D}_{21} & \underline{D}_{22} \end{pmatrix}$. Write $\underline{D}_{22.1} = \underline{D}_{22} - \underline{D}_{21}\,\underline{D}_{22}^{-1}\underline{D}_{12}$. Then the posterior distribution of $\underline{\beta}$ given $\underline{Y} = \underline{y}$ is as follows:

Theorem 2.3.1. Conditional on $\underline{Y} = \underline{y}$, $R_1 = r_1$ and $R_2 = r_2$,

$$\underline{\beta} \sim N\Big(\begin{bmatrix} \hat{\underline{\beta}}_1 - r_1 \underline{C}_{11}^{-1}\underline{C}_{12}(r_1\underline{C}_{22.1} + r_2\underline{D}_{22.1})^{-1}\underline{C}_{22.1}\tilde{\underline{\beta}}_2 \\ r_1(r_1\underline{C}_{22.1} + r_2\underline{D}_{22.1})^{-1}\underline{C}_{22.1}\tilde{\underline{\beta}}_2 \end{bmatrix},$$
$$\begin{bmatrix} r_1\underline{C}_{11} & r_1\underline{C}_{12} \\ r_1\underline{C}_{21} & r_1\underline{C}_{22} + r_2\underline{D}_{22.1} \end{bmatrix}^{-1} \Big). \tag{2.66}$$

Also, the joint posterior distribution of $R_1 = r_1$ and $R_2 = r_2$ given $\underline{Y} = \underline{y}$ is given by

$$f(r_1, r_2|\underline{y}) \propto r_1^{\frac{n-p+f_1}{2}-1} \times r_2^{\frac{p_2+f_2}{2}-1} |\underline{C}_{22.1} + \frac{r_2}{r_1}\underline{D}_{22.1}|^{-\frac{1}{2}}$$
$$\times \exp(-\frac{1}{2}r_1 SSE - \frac{1}{2}a_1 r_1) \exp(-\frac{1}{2}r_2 \underline{\tilde{\beta}}_2^T (\underline{D}_{22.1}^{-1} + \frac{r_2}{r_1}\underline{C}_{22.1}^{-1})^{-1}\underline{\tilde{\beta}}_2 - \frac{1}{2}a_2 r_2) \tag{2.67}$$

Proof of Theorem 2.3.1. Writing $\underline{C} = \underline{X}^T \underline{V}^{-1} \underline{X}$, the joint *pdf* of $\underline{\gamma}$, $\underline{\beta}$, $\underline{\nu}_1$, R_1 and R_2 is

$$\begin{aligned} & f(\underline{\gamma}, \underline{\beta}, \underline{\nu}, r_1, r_2) \\ \propto\ & r_1^{n/2} \exp[-\frac{1}{2} r_1 \{(\underline{\gamma} - \underline{X}\underline{\tilde{\beta}})^T \underline{V}^{-1} (\underline{\gamma} - \underline{X}\underline{\tilde{\beta}}) + (\underline{\tilde{\beta}} - \underline{\beta})^T \underline{C} (\underline{\tilde{\beta}} - \underline{\beta})\}] \\ & \times r_2^{p/2} \exp[-\frac{1}{2} r_2 (\underline{\beta} - \underline{\nu})^T \underline{D} (\underline{\beta} - \underline{\nu})] \\ & \times r_1^{\frac{f_1}{2}-1} \exp(-\frac{1}{2} a_1 r_1) r_2^{\frac{f_2}{2}-1} \exp(-\frac{1}{2} a_2 r_2). \end{aligned} \tag{2.68}$$

Since $\underline{\nu}^T = (\underline{\nu}_1^T, \underline{0}^T)$, writing $\underline{D} = \begin{pmatrix} \underline{D}_{11} & \underline{D}_{12} \\ \underline{D}_{21} & \underline{D}_{22} \end{pmatrix}$,

$$\begin{aligned} & (\underline{\beta} - \underline{\nu})^T \underline{D} (\underline{\beta} - \underline{\nu}) \\ = & (\underline{\beta}_1 - \underline{\nu}_1)^T \underline{D}_{11} (\underline{\beta}_1 - \underline{\nu}_1) + \underline{\beta}_2^T \underline{D}_{22} \underline{\beta}_2 + 2(\underline{\beta}_1 - \underline{\nu}_1)^T \underline{D}_{12} \underline{\beta}_2 \\ = & (\underline{\nu}_1 - \underline{\beta}_1 - \underline{D}_{11}^{-1} \underline{D}_{12} \underline{\beta}_2)^T \underline{D}_{11} (\underline{\nu}_1 - \underline{\beta}_1 - \underline{D}_{11}^{-1} \underline{D}_{12} \underline{\beta}_2) + \underline{\beta}_2^T \underline{D}_{22.1} \underline{\beta}_2. \end{aligned} \tag{2.69}$$

Now, integrating with respect to $\underline{\nu}_1$, it follows from Eq.(2.68) that

$$\begin{aligned} & f(\underline{\gamma}, \underline{\beta}, r_1, r_2) \\ \propto\ & r_1^{n/2} \exp[-\frac{1}{2} r_1 \{SSE + (\underline{\tilde{\beta}} - \underline{\beta})^T \underline{C} (\underline{\tilde{\beta}} - \underline{\beta})\}] \\ \times\ & r_2^{p_2/2} \exp(-\frac{1}{2} r_2 \underline{\beta}_2^T \underline{D}_{22.1} \underline{\beta}_2) r_1^{\frac{1}{2}f_1 - 1} \exp(-\frac{1}{2} a_1 r_1) r_2^{-\frac{1}{2}f_2 - 1} \exp(-\frac{1}{2} a_2 r_2). \end{aligned} \tag{2.70}$$

Now, using $\underline{\tilde{\beta}}_1 = \underline{\hat{\beta}}_1 - \underline{C}_{11}^{-1} \underline{C}_{12} \underline{\tilde{\beta}}_2$, it can be shown that

$$\begin{aligned} & r_1 (\underline{\tilde{\beta}} - \underline{\beta})^T \underline{C} (\underline{\tilde{\beta}} - \underline{\beta}) + r_2 \underline{\beta}_2^T \underline{D}_{22.1} \underline{\beta}_2 \\ = & r_1 [(\underline{\beta}_1 - \underline{\hat{\beta}}_1 + \underline{C}_{11}^{-1} \underline{C}_{12} \underline{\tilde{\beta}}_2)^T \underline{C}_{11} (\underline{\beta}_1 - \underline{\hat{\beta}}_1 + \underline{C}_{11}^{-1} \underline{C}_{12} \underline{\tilde{\beta}}_2) + (\underline{\beta}_2 - \underline{\tilde{\beta}}_2)^T \underline{C}_{22} (\underline{\beta}_2 - \underline{\tilde{\beta}}_2) \\ & + 2(\underline{\beta}_2 - \underline{\tilde{\beta}}_2)^T \underline{C}_{21} (\underline{\beta}_1 - \underline{\hat{\beta}}_1 + \underline{C}_{11}^{-1} \underline{C}_{12} \underline{\tilde{\beta}}_2)] + r_2 \underline{\beta}_2^T \underline{D}_{22.1} \underline{\beta}_2 \\ = & r_1 [(\underline{\beta}_1 - \underline{\hat{\beta}}_1)^T \underline{C}_{11} (\underline{\beta}_1 - \underline{\hat{\beta}}_1) + \underline{\tilde{\beta}}_2^T \underline{C}_{22.1} \underline{\tilde{\beta}}_2 - 2 \underline{\beta}_2^T \underline{C}_{22.1} \underline{\tilde{\beta}}_2 \\ & + 2 \underline{\beta}_2^T \underline{C}_{21} (\underline{\beta}_1 - \underline{\hat{\beta}}_1)] + \underline{\beta}_2^T (r_1 \underline{C}_{22} + r_2 \underline{D}_{22.1}) \underline{\beta}_2 \\ = & r_1 [(\underline{\beta}_1 - \underline{\hat{\beta}}_1 + \underline{A} \underline{\tilde{\beta}}_2)^T \underline{C}_{11} (\underline{\beta}_1 - \underline{\hat{\beta}}_1 + \underline{A} \underline{\tilde{\beta}}_2)] \\ & + (\underline{\beta}_2 - \underline{K} \underline{\tilde{\beta}}_2)^T (r_1 \underline{C}_{22} + r_2 \underline{D}_{22.1}) (\underline{\beta}_2 - \underline{K} \underline{\tilde{\beta}}_2) \\ & + 2 r_1 (\underline{\beta}_2 - \underline{K} \underline{\tilde{\beta}}_2)^T \underline{C}_{21} (\underline{\beta}_1 - \underline{\hat{\beta}}_1 + \underline{A} \underline{\tilde{\beta}}_2) \\ & + \underline{\tilde{\beta}}_2^T (r_1 \underline{C}_{22.1} - r_1 \underline{A}^T \underline{C}_{11} \underline{A} - \underline{K}^T (r_1 \underline{C}_{22} + r_2 \underline{D}_{22.1}) \underline{K} + 2 r_1 \underline{K}^T \underline{C}_{21} \underline{A}) \underline{\tilde{\beta}}_2, \end{aligned} \tag{2.71}$$

where $\underline{A}$ and $\underline{K}$ satisfy

$$2r_1(\underline{A}^T \underline{C}_{11} - \underline{K}^T \underline{C}_{21}) = \underline{0}; \tag{2.72}$$

$$-2r_1 \underline{C}_{22.1} = -2(r_1 \underline{C}_{22} + r_2 \underline{D}_{22.1})\underline{K} + 2r_1 \underline{C}_{21} \underline{A}. \tag{2.73}$$

Solving (2.72) and (2.73) one gets

$$\underline{K} = r_1(r_1 \underline{C}_{22.1} + r_2 \underline{D}_{22.1})^{-1} \underline{C}_{22.1}, \tag{2.74}$$

$$\underline{A} = \underline{C}_{11}^{-1} \underline{C}_{12} \underline{K}. \tag{2.75}$$

Using (2.74) and (2.75), part of the last term in Eq.(2.71) can be expressed as:

$$\begin{aligned}
& r_1(\underline{C}_{22.1} - \underline{A}^T \underline{C}_{11} \underline{A} + 2\underline{K}^T \underline{C}_{21} \underline{A}) - \underline{K}^T (r_1 \underline{C}_{22} + r_2 \underline{D}_{22.1}) \underline{K} \\
= \; & r_1 \underline{C}_{22.1} - r_1 \underline{C}_{22.1} (\underline{C}_{22.1} + \frac{r_2}{r_1} \underline{D}_{22.1})^{-1} \underline{C}_{22.1} \\
= \; & r_1 \underline{C}_{22.1} - r_1 (\underline{C}_{22.1} + \frac{r_2}{r_1} \underline{D}_{22.1} - \frac{r_2}{r_1} \underline{D}_{22.1}) (\underline{C}_{22.1} + \frac{r_2}{r_1} \underline{D}_{22.1})^{-1} \underline{C}_{22.1} \\
= \; & r_2 \underline{D}_{22.1} (\underline{C}_{22.1} + \frac{r_2}{r_1} \underline{D}_{22.1})^{-1} \underline{C}_{22.1} \\
= \; & r_2 \underline{D}_{22.1} (\underline{C}_{22.1} (\underline{D}_{22.1}^{-1} + \frac{r_2}{r_1} \underline{C}_{22.1}^{-1}) \underline{D}_{22.1})^{-1} \underline{C}_{22.1} \\
= \; & r_2 (\underline{D}_{22.1}^{-1} + \frac{r_2}{r_1} \underline{C}_{22.1}^{-1})^{-1}.
\end{aligned} \tag{2.76}$$

It follows (2.70), (2.71), (2.74), (2.75) and (2.76) that the posterior distribution of $\underline{\beta}$ given $\underline{y}$, r_1 and r_2 is given by (2.66).

Again, following (2.70), (2.71), (2.74), (2.75), (2.76) and

$$\begin{vmatrix} r_1 \underline{C}_{11} & r_1 \underline{C}_{12} \\ r_1 \underline{C}_{21} & r_1 \underline{C}_{22} + r_2 \underline{D}_{22.1} \end{vmatrix}^{1/2}$$

$$\begin{aligned}
&= |r_1 \underline{C}_{11}|^{1/2} |r_1 \underline{C}_{22.1} + r_2 \underline{D}_{22.1}|^{1/2} \\
&= r_1^{p/2} |\underline{C}_{11}| |\underline{C}_{22.1} + \frac{r_2}{r_1} \underline{D}_{22.1}|,
\end{aligned} \tag{2.77}$$

it shows the joint *pdf* of $\underline{Y}$, R_1, and R_2 as:

$$f(\underline{y}, r_1, r_2)$$

$$
\begin{aligned}
&\propto r_1^{\frac{1}{2}n} r_2^{\frac{1}{2}p_2} r_1^{-\frac{1}{2}p} |\underline{C}_{22.1} + \frac{r_2}{r_1}\underline{D}_{22.1}|^{-\frac{1}{2}} \\
&\quad \times \exp[-\frac{1}{2}r_1 SSE - \frac{1}{2}r_2 \tilde{\underline{\beta}}_2^T (\underline{D}_{22.1}^{-1} + \frac{r_2}{r_1}\underline{C}_{22.1}^{-1})^{-1} \tilde{\underline{\beta}}_2] \\
&\quad \times \exp(-\frac{1}{2}r_1 a_1) r_1^{\frac{1}{2}f_1 - 1} \exp(-\frac{1}{2}r_2 a_2) r_2^{\frac{1}{2}f_2 - 1} \\
&\propto r_1^{\frac{1}{2}(n-p+f_1)-1} r_2^{\frac{1}{2}(p_2+f_2)-1} |\underline{C}_{22.1} + \frac{r_2}{r_1}\underline{D}_{22.1}|^{-\frac{1}{2}} \\
&\quad \times \exp[-\frac{1}{2}r_1 (SSE + a_1)] \\
&\times \quad \exp[-\frac{1}{2}r_2 \{\tilde{\underline{\beta}}_2^T (\underline{D}_{22.1}^{-1} + \frac{r_2}{r_1}\underline{C}_{22.1})^{-1} \tilde{\underline{\beta}}_2 + a_2\}]. \qquad (2.78)
\end{aligned}
$$

This leads to (2.67). The proof of Theorem 2.3.1 is completed.

It is possible do draw inference about $\underline{\beta}$ or $\underline{X}\underline{\beta}$ based on the above posterior distribution. One can obtain expression for $E(\underline{X}\underline{\beta}|\underline{Y} = \underline{y})$ and $V(\underline{X}\underline{\beta}|\underline{Y} = \underline{y})$ which will lead to point estimators as well as credible sets for $\underline{X}\underline{\beta}$. In the special case of g-priors, the above posterior distribution, however, simplifies considerably.

To see this note that for the g-prior of Section 2.2, $D = \underline{X}^T \underline{V}^{-1} \underline{X}$. Then, writing $B = R_2/(R_1 + R_2)$, $b = r_2/(r_1 + r_2)$, and recalling that $\tilde{\underline{\beta}}_1 = \hat{\underline{\beta}}_1 - \underline{C}_{11}^{-1}\underline{C}_{12}\tilde{\underline{\beta}}_2$, it follows from (2.66) that conditional on $\underline{Y} = \underline{y}$, $R_1 = r_1$ and $R_2 = r_2$,

$$
\underline{\beta} \sim N(\begin{bmatrix} \hat{\underline{\beta}}_1 + (1-b)(\tilde{\underline{\beta}}_1 - \hat{\underline{\beta}}_1) \\ (1-b)\underline{\beta}_2 \end{bmatrix}, r_1^{-1} \begin{bmatrix} \underline{C}_{11} & \underline{C}_{12} \\ \underline{C}_{21} & \underline{C}_{22} + (b/1-b)\underline{C}_{22.1} \end{bmatrix}^{-1}). \qquad (2.79)
$$

Using the usual inversion formula for partitioned matrices (see e.g., Exercise 2.7, p.33 of Rao, 1973), one gets,

$$
\begin{aligned}
&\begin{bmatrix} \underline{C}_{11} & \underline{C}_{12} \\ \underline{C}_{21} & \underline{C}_{22} + (b/1-b)\underline{C}_{22.1} \end{bmatrix}^{-1} \\
&= \begin{bmatrix} \underline{C}_{11}^{-1} + (1-b)\underline{C}_{11}^{-1}\underline{C}_{12}\underline{C}_{22.1}^{-1}\underline{C}_{21}\underline{C}_{11}^{-1} & -(1-b)\underline{C}_{11}^{-1}\underline{C}_{12}\underline{C}_{22.1}^{-1} \\ -(1-b)\underline{C}_{22.1}^{-1}\underline{C}_{21}\underline{C}_{11}^{-1} & (1-b)\underline{C}_{22.1}^{-1} \end{bmatrix} \qquad (2.80)
\end{aligned}
$$

Also, it follows from (2.67) that under the g-prior, the joint posterior distribution of R_1 and R_2 given $\underline{Y} = \underline{y}$ is

$$
\begin{aligned}
&f(r_1, r_2|\underline{y}) \\
&\propto r_1^{\frac{1}{2}(n-p+f_1)-1} r_2^{\frac{1}{2}(p_2+f_2)-1} (\frac{r_1}{r_1+r_2})^{\frac{1}{2}p_2} \\
&\quad \times \exp[-\frac{1}{2}r_1 (SSE + a_1 + \frac{r_2}{r_1+r_2} SSH) - \frac{1}{2}a_2 r_2]. \qquad (2.81)
\end{aligned}
$$

Since $B = R_2/(R_1 + R_2)$, the joint posterior distribution of R_1 and B given $\underline{Y} = \underline{y}$ is

$$f(r_1, b|\underline{y}) \propto r_1^{\frac{1}{2}(n-p_1+f_1+f_2)-1} b^{\frac{1}{2}p_2-2} \left(\frac{b}{1-b}\right)^{\frac{1}{2}f_2+1} \times \exp\left[-\frac{1}{2}r_1(SSE + a_1 + bSSH + a_2 b(1-b)^{-1})\right]. \quad (2.82)$$

It follows from (2.82) that the posterior distribution of B given $\underline{Y} = \underline{y}$ is

$$f(b|\underline{y}) \propto b^{\frac{1}{2}p_2-2}\left(\frac{b}{1-b}\right)^{\frac{1}{2}f_2+1}(SSE + a_1 + bSSH + a_2 b(1-b)^{-1})^{-\frac{1}{2}(n-p_1+f_1+f_2)}. \quad (2.83)$$

Now, from (2.79), since $\underline{\tilde{\beta}}_1 = \underline{\hat{\beta}}_1 - \underline{C}_{11}^{-1}\underline{C}_{12}\underline{\tilde{\beta}}_2$

$$E(\underline{X}\,\underline{\beta}|\underline{Y}) = \underline{X}_1\,\underline{\hat{\beta}}_1 + (\underline{X}_2 - \underline{X}_1\,\underline{C}_{11}^{-1}\underline{C}_{12})(1 - E(B|\underline{Y}))\underline{\tilde{\beta}}_2. \quad (2.84)$$

Also,

$$\begin{aligned} & V[\,E(\underline{X}\,\underline{\beta}|\underline{Y}, B, R_1)|\underline{Y}\,] = V[\,\underline{X}_1\,\underline{\hat{\beta}}_1 + (1-B)(\underline{X}_2 - \underline{X}_1\,\underline{C}_{11}^{-1}\underline{C}_{12})\underline{\tilde{\beta}}_2|\underline{Y}\,] \\ = \; & V(B|\underline{Y})(\underline{X}_2 - \underline{X}_1\underline{C}_{11}^{-1}\underline{C}_{12})\underline{\tilde{\beta}}_2\underline{\tilde{\beta}}_2^T(\underline{X}_2 - \underline{X}_1\underline{C}_{11}^{-1}\underline{C}_{12})^T. \end{aligned} \quad (2.85)$$

$$E[\,V(\underline{X}\,\underline{\beta}|\underline{Y}, B, R_1)|\underline{Y}\,] = \underline{X}\begin{bmatrix} \underline{S}_{11} & \underline{S}_{12} \\ \underline{S}_{21} & \underline{S}_{22} \end{bmatrix}\underline{X}^T, \quad (2.86)$$

where

$$\begin{aligned} \underline{S}_{11} &= \underline{C}_{11}^{-1}E(R_1^{-1}|\underline{Y}) + E(R_1^{-1}(1-B)|\underline{Y})\underline{C}_{11}^{-1}\underline{C}_{12}\underline{C}_{22.1}^{-1}\underline{C}_{21}\underline{C}_{11}^{-1} \\ \underline{S}_{12} &= -E(R_1^{-1}(1-B)|\underline{Y})\underline{C}_{11}^{-1}\underline{C}_{12}\underline{C}_{22.1}^{-1} \\ \underline{S}_{21} &= -E(R_1^{-1}(1-B)|\underline{Y})\underline{C}_{22.1}^{-1}\underline{C}_{21}\underline{C}_{11}^{-1} \\ \underline{S}_{22} &= E(R_1^{-1}(1-B)|\underline{Y})\underline{C}_{22.1}^{-1} \end{aligned}$$

Using the formula

$$V(\underline{X}\,\underline{\beta}|\underline{Y}) = E[\,V(\underline{X}\,\underline{\beta}|\underline{Y}, B, R_1)|\underline{Y}\,] + V[\,E(\underline{X}\,\underline{\beta}|\underline{Y}, B, R_1)|\underline{Y}\,], \quad (2.87)$$

one can find $V(\underline{X}\,\underline{\beta}|\underline{Y})$ from (2.85) to (2.87).

It follows from (2.82) that

$$E(R_1^{-1}|b,\underline{y}) = \frac{SSE + a_1 + b\,SSE + a_2 b(1-b)^{-1}}{n - p_1 + f_1 + f_2 - 2}. \tag{2.88}$$

The calculations for (2.86) can now be completed from (2.83) and (2.88).

To compare the HB estimator given in (2.84) with the EB estimator of Section 2.2, consider the case when $a_2 = 0$, $f_2 = -2$ and $a_1 = 0$. Thus, the prior distribution on R_1 as well as R_2 are improper. Then the posterior distribution given in (2.66) simplifies to

$$f(b|\underline{y}) \propto b^{\frac{1}{2}p_2 - 2}(SSE + b\,SSH)^{-\frac{1}{2}(n-p_1+f_1-2)}. \tag{2.89}$$

From (2.89), one gets

$$E(B|\underline{Y}) = \frac{\int_0^1 b^{\frac{1}{2}p_2-1}(1+\frac{bF}{n-p+2})^{-\frac{1}{2}(n-p_1+f_1-2)}db}{\int_0^1 b^{\frac{1}{2}p_2-2}(1+\frac{bF}{n-p+2})^{-\frac{1}{2}(n-p_1+f_1-2)}db}. \tag{2.90}$$

This can be written in the form $\phi(F)/F$ mentioned in Section 2.2. Also, instead of taking the posterior mean of B, suppose one considers the posterior mode. The posterior mode or the Type II ML-estimator of B as obtained from (2.89) is

$$\hat{B}_{MO} = min\left(\frac{(n-p+2)(p_2-4)}{(n-p+f_1+2)F}, 1\right), \tag{2.91}$$

if $p_2 > 5$. Thus the estimator $\underline{X}_1[\,\underline{\hat{\beta}}_1 + (1-\hat{B}_{MO})(\underline{\tilde{\beta}}_1 - \underline{\hat{\beta}}_1)\,] + \underline{X}_2(1-\hat{B}_{MO})\underline{\tilde{\beta}}_2$ of $\underline{X}\,\underline{\beta}$ is a special member of the class of estimator $\underline{e}_\phi^+$ of $\underline{X}\,\underline{\beta}$.

2.4 Numerical Examples

This section illustrates numerically some of the estimation results derived in the previous sections. Monte-Carlo experiments are conducted to assess the frequentist as well as the Bayes risks of the different estimators. In addition, real data examples were used to illustrate the applicability of the proposed methods.

Two situations were considered in the Monte-Carlo experiments. The first involved a 2^3 factorial experiment replicated twice. Suppose experimenter assumes a response surface model of the form.

$$E(Y_{ij}) = \gamma_0 + \gamma_1\zeta_{1i} + \gamma_2\zeta_{2i} + \gamma_3\zeta_{3i} + \gamma_{12}\zeta_{1i}\zeta_{2i} + \gamma_{13}\zeta_{1i}\zeta_{3i} + \gamma_{23}\zeta_{2i}\zeta_{3i} + \gamma_{123}\zeta_{1i}\zeta_{2i}\zeta_{3i},$$

where $i = 1, \cdots, 8$ (treatment) and $j = 1, 2$ (replication). $\varsigma_{1i} = 1$ for $i = 1, 2, 3, 4$, corresponding to the high level, and $\varsigma_{1i} = -1$ for $i = 5, 6, 7, 8$, corresponding to the low level, of the variable (factor) ς_1. Also $\varsigma_{2i} = 1$ for $i = 1, 2, 5, 6$, $\varsigma_{2i} = -1$ for $i = 3, 4, 7, 8$, $\varsigma_{3i} = 1$ for $i = 1, 3, 5, 7$ and $\varsigma_{3i} = -1$ for $i = 2, 4, 6, 8$. For $\underline{\beta}_1^T = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)$ and $\underline{\beta}_2^T = (\gamma_{12}, \gamma_{13}, \gamma_{23}, \gamma_{123})$, then, $p_1 = p_2 = 4$ and $n = 16$. The design matrix $\underline{X}$ can be partitioned as $\underline{X} = (\underline{X}_1, \underline{X}_2)$, where

$$\underline{X}_1^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \end{bmatrix}$$

$$\underline{X}_2^T = \begin{bmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \end{bmatrix}$$

Note that $\underline{X}^T\underline{X} = 16 \times \underline{I}_8$. It is assumed that $\underline{V}_n = \underline{I}_{16}$ and $\underline{Q} = \underline{I}_{16}$ here.

The second example a 3^2 factorial experiment, involved two factors each at three levels. This is taken from Myers (1976). It assumes a respone surface model of the form

$$E(\underline{Y}_i) = \gamma_0 + \gamma_1\varsigma_{1i} + \gamma_2\varsigma_{2i} + \gamma_{11}\varsigma_{1i}^2 + \gamma_{22}\varsigma_{2i}^2 + \gamma_{12}\varsigma_{1i}\varsigma_{2i},$$

$i = 1, \cdots, 9$, where $\varsigma_{1i} = -1$ for $i = 1, 2, 3$, corresponding to the low level, $\varsigma_{1i} = 0$ for $i = 4, 5, 6$, corresponding to the central lavel, and $\varsigma_{1i} = 1$ for $i = 7, 8, 9$, corresponding to the high level, of the variable ς_1. Also, $\varsigma_{2i} = -1$ for $i = 1, 4, 7$; $\varsigma_{2i} = 0$ for $i = 2, 5$ and 8, and $\varsigma_{2i} = 1$ for $i = 3, 6$ and 9. Here $\underline{\beta}_1^T = (\gamma_0, \gamma_1, \gamma_2)$, comprises the general effect and the two main effects, while $\underline{\beta}_2^T = (\gamma_{11}, \gamma_{22}, \gamma_{12})$, comprises the second order interactions. Thus $p_1 = p_2 = 3$ and $n = 9$. One can partition the design matrix $\underline{X}$ as $\underline{X} = (\underline{X}_1, \underline{X}_2)$, where

$$\underline{X}_1^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & 0 & 0 & 0 & 1 & 1 & 1 \\ -1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 & 1 \end{bmatrix}$$

$$\underline{X}_2^T = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix}.$$

Table 2.1: The estimated frequentist risk results for 3^2 factorial experiment.

The estimator of $\underline{X}\underline{\beta}$	$\underline{\beta}$	
	(3 0 1 1 2 3)	(3 0 1 0 0 0)
$\underline{X}\,\widetilde{\underline{\beta}}$	5.99478	6.03603
$\underline{X}_1\,\widehat{\underline{\beta}}_1$	48.9807	2.97617
$\underline{e}_{EB}(\underline{Y})$	5.99216	5.32031
$\underline{e}^+_{EB}(\underline{Y})$	5.99216	5.16984
$\underline{e}_{HB}(\underline{Y})$	5.99877	5.33225
$\underline{e}_{PTE.10}(\underline{Y})$	7.81844	3.60394
$\underline{e}_{MEB.10}(\underline{Y})$	7.79382	3.57963
$\underline{e}^+_{MEB.10}(\underline{Y})$	7.79382	3.57963
$\underline{e}_{PTE.05}(\underline{Y})$	13.8542	3.0364
$\underline{e}_{MEB.05}(\underline{Y})$	13.8169	3.29718
$\underline{e}^+_{MEB.05}(\underline{Y})$	13.8169	3.29718
$\underline{e}_{PTE.01}(\underline{Y})$	34.2084	3.10328
$\underline{e}_{MEB.01}(\underline{Y})$	34.1966	3.10220
$\underline{e}^+_{MEB.01}(\underline{Y})$	34.1966	3.10220

Also, $\underline{V}_n = \underline{I}_9$ and $\underline{Q} = \underline{I}_9$ are assumed in this example. Here $\underline{C}_{11} = Diag(9,6,6)$, $\underline{C}_{22} = \begin{bmatrix} 6 & 4 & 0 \\ 4 & 6 & 0 \\ 0 & 0 & 4 \end{bmatrix}$ and $\underline{C}_{12} = \begin{bmatrix} 6 & 6 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, from which, one computes $\underline{C}_{22.1} = Diag(2,2,4)$.

For a given $\underline{\beta}$, a random vector $\underline{Y}_n$ was generated from $N(\underline{X}\,\underline{\beta}, \underline{I}_n)$ in each replication. The sampling procedure was repeated 1,000 times. For estimating $\underline{X}\,\underline{\beta}$, the estimated frequentist risks, i.e., the average squared error losses of the least squares estimators (LSE, full model and $RLSE$, reduced model), the empirical Bayes estimator ($\underline{e}_{EB}(\underline{Y})$), the positive part empirical Bayes estimator ($\underline{e}^+_{EB}(\underline{Y})$), the hierarchical Bayes estimator ($\underline{e}_{HB}(\underline{Y})$, as given in (2.90) with $f_1 = 2$), the preliminary test estimators ($\underline{e}_{PTE\alpha}(\underline{Y})$'s, α being the significant level, = .10, .05, .01), the modified empirical Bayes estimators ($\underline{e}_{MEB\alpha}(\underline{Y})$'s) and the positive part modified empirical Bayes estimators ($\underline{e}^+_{MEB\alpha}(\underline{Y})$'s) are compared. The results for the 3^2 experiment are given in Table 2.1, while the results for the 2^3 experiment are given in Table 2.2.

Table 2.2: The estimated frequentist risk results for 2^3 factorial experiment replicated two times.

The estimator of $\underline{X}\underline{\beta}$	$\underline{\beta}$	
	(-1 4 0 3 2 1 5 -2)	(-1 4 0 3 0 0 0 0)
$\underline{X}\underline{\tilde{\beta}}$	7.77952	7.92404
$\underline{X}_1\underline{\hat{\beta}}_1$	547.88700	3.93829
$\underline{e}_{EB}(\underline{Y})$	7.77666	6.17508
$\underline{e}^+_{EB}(\underline{Y})$	7.77666	5.83407
$\underline{e}_{HB}(\underline{Y})$	7.78255	6.18972
$\underline{e}_{PTE.10}(\underline{Y})$	7.77952	4.81253
$\underline{e}_{MEB.10}(\underline{Y})$	7.77666	4.65800
$\underline{e}^+_{MEB.10}(\underline{Y})$	7.77666	4.65800
$\underline{e}_{PTE.05}(\underline{Y})$	7.77952	4.47572
$\underline{e}_{MEB.05}(\underline{Y})$	7.77666	4.39852
$\underline{e}^+_{MEB.05}(\underline{Y})$	7.77666	4.39852
$\underline{e}_{PTE.01}(\underline{Y})$	7.77952	4.05713
$\underline{e}_{MEB.01}(\underline{Y})$	7.77666	4.04630
$\underline{e}^+_{MEB.01}(\underline{Y})$	7.77666	4.04630

* Note: For $\underline{\beta} = (-1\,4\,0\,3\,2\,1\,5\,-2)$, during the 1,000 replications, the null hypothesis, $\underline{\beta}_2 = \underline{0}$, was rejected in every replication for all α =.10, .05 and .01. Hence, $\underline{e}_{PTE\alpha}(\underline{Y}) = LSE$, $\underline{e}_{MEB\alpha}(\underline{Y}) = \underline{e}_{EB}(\underline{Y})$ and $\underline{e}^+_{MEB\alpha}(\underline{Y}) = \underline{e}^+_{EB}(\underline{Y})$.

When $\underline{\beta}_2 = \underline{0}$, $\underline{X}_1\hat{\underline{\beta}}_1$, the LSE of $\underline{X}\underline{\beta}$ under the reduced model, is the best estimator of $\underline{X}\underline{\beta}$. However, substantial improvement in the estimated frequentist risk of $\underline{e}^+_{EB}$ over $\underline{X}\tilde{\underline{\beta}}$, the full model LSE, was noticed. A similar phenomenon occured with the hierarchical Bayes estimator. Other HB estimators with $f_2 = 1$ and 4 were tried. The results are not reported, because they are very similar to the case $f_2 = 2$. Note also that the estimated frequentist risks of the PTE's, the MEB's and the MEB^+'s are all decreasing with increase in the significant level. This is because the larger the critical value, the less is the chance to reject $H_0 : \underline{\beta}_2 = \underline{0}$.

When $\underline{\beta}_2 \neq \underline{0}$, $\underline{e}_{EB}$ or $\underline{e}^+_{EB}$ are the estimators having the smallest frequentist risks. Admittedly, the risk improvement is not significant over the full model LSE. However, there is considerable risk improvement over the reduced model LSE, as well as the different PTE's, and the modified $\underline{e}_{EB}$'s or $\underline{e}^+_{EB}$'s. In this case, the ordering of the estimated frequentist risks of the PTE's, the MEB's and the MEB^+'s is reversed. This can be accounted for by the same argument given at the end of the previous paragragh.

Next the Bayes risks ware considered. For each replication, $\underline{\beta}$ was generated from $N(\underline{\nu}, \tau^2(\underline{X}^T\underline{X})^{-1})$ for a given $\underline{\nu}$ and τ^2. For a generated $\underline{\beta}$ and a given σ^2, a random vector $\underline{Y}_n$ was then generated from $N(\underline{X}\underline{\beta}, \sigma^2\underline{I}_n)$. Once again, the sampling procedure was repeated 1,000 times. The average, over 1,000 replications, squared error losses of the different estimators of $\underline{X}\underline{\beta}$ ware compared. Tables 2.3 and 2.4 display the estimated Bayes risks of the different estimators.

When $\tau^2 < \sigma^2$, i.e. the prior variability is smaller than the sampling variability, the estimated Bayes risk of the different estimator maintain the same order as the corresponding estimated frequentist risks. The case $\underline{\nu}_2 = \underline{0}(\underline{\nu}_2 \neq \underline{0})$ corresponds to the case $\underline{\beta}_2 = \underline{0}$ $(\underline{\beta}_2 \neq \underline{0})$. However, when $\tau^2 > \sigma^2$, even though $\underline{\nu}_2 = \underline{0}$, the reduced model LSE, the PTE's and the modified $\underline{e}_{EB}$'s or $\underline{e}^+_{EB}$'s all perform poorly as compared to the full model LSE, $\underline{e}_{EB}$ or $\underline{e}^+_{EB}$. The reason is that increased

Table 2.3: The estimated Bayes risk results for 3^2 factorial experiment.

The	$\underline{\nu}$	(2 3 4 1 1 1)			(2 3 4 0 0 0)		
estimator	σ^2	1			1		
of $\underline{X}\underline{\beta}$	τ^2	4	1	.25	4	1	.25
$\underline{X}\hat{\underline{\beta}}$		5.95497	5.95497	5.95497	5.95497	5.95497	5.95497
$\underline{X}_1\hat{\underline{\beta}}_1$		22.8549	13.9354	11.7289	15.0423	6.02910	3.77581
$\underline{e}_{EB}(\underline{Y})$		5.89124	5.87080	5.88644	5.82264	5.60552	5.40601
$\underline{e}^+_{EB}(\underline{Y})$		5.87863	5.86034	5.87390	5.79278	5.53087	5.27738
$\underline{e}_{HB}(\underline{Y})$		5.80484	5.73912	5.72836	5.69700	5.43529	5.26267
$\underline{e}_{PTE.10}(\underline{Y})$		9.04118	9.47173	9.25189	8.87426	6.06646	4.31691
$\underline{e}_{MEB.10}(\underline{Y})$		8.97500	9.40509	9.18729	8.81221	6.02341	4.28660
$\underline{e}^+_{MEB.10}(\underline{Y})$		8.97500	9.40509	9.18729	8.81221	6.02341	4.28660
$\underline{e}_{PTE.05}(\underline{Y})$		12.4347	11.0010	10.2346	10.5683	6.06056	4.07614
$\underline{e}_{MEB.05}(\underline{Y})$		12.3974	10.9697	10.2076	10.5347	6.04193	4.06642
$\underline{e}^+_{MEB.05}(\underline{Y})$		12.3974	10.9796	10.2076	10.5347	6.04193	4.06642
$\underline{e}_{PTE.01}(\underline{Y})$		19.0021	13.0864	11.3694	13.6495	6.03266	3.84518
$\underline{e}_{MEB.01}(\underline{Y})$		18.9955	13.0824	11.3662	13.6454	6.03266	3.84450
$\underline{e}^+_{MEB.01}(\underline{Y})$		18.9955	13.0824	11.3662	13.6454	6.14803	3.84450

prior variability overshadows the credibility of the prior mean. When $\tau^2 = \sigma^2$, the EB or the postive part EB estimator dominate both the LSE or the $RLSE$. The domination is much more pronounced in the case when $\underline{\nu}_2$ indeed equals to $\underline{0}$.

Next, we investigate how these estimators perfrom on some real data. Professor James Barret and Terril Nell of the University of Florida investigated the effects of growth retardant chemicals at various rates of application on heights of chrysanthemum plants (personal communication). Their experiment was a randomized complete block design with five blocks. Two chemicals were each applied at six rates, given 2×6 factorial treatment structure. The 12 treatments were applied to two plots in each block. The estimation methods discussed in this paper will be illustrated using data from two of the five blocks, block one and block five. Seperate illustrations are presented for blocks 1 and 5 because quite different assessments of chemical$\times$rate interaction are revealed in the two blocks. Data are presented in

Table 2.4: The estimated Bayes risk results for 2^3 factorial experiment replicated two times.

The estimator of $\underline{X}\underline{\beta}$	$\underline{\nu}$	(-4 6 8 3 5 2 7 4)			(-4 6 8 3 0 0 0 0)		
	σ^2	4			4		
	τ^2	9	4	1	9	4	1
$\underline{X}\tilde{\underline{\beta}}$		31.7548	31.7548	31.7548	31.7548	31.7548	31.7548
$\underline{X}_1\hat{\underline{\beta}}_1$		1556.05	1536.09	1524.16	52.3833	32.3134	20.2714
$\underline{e}_{EB}(\underline{Y})$		31.7530	31.7526	31.7522	30.4188	29.5948	27.1787
$\underline{e}^+_{EB}(\underline{Y})$		31.7530	31.7526	31.7522	29.6207	27.8240	25.1256
$\underline{e}_{HB}(\underline{Y})$		31.7892	31.7895	31.7896	28.9510	27.3352	25.4736
$\underline{e}_{PTE.10}(\underline{Y})$		31.7548	31.7548	31.7548	37.1159	32.1043	23.1223
$\underline{e}_{MEB.10}(\underline{Y})$		31.7530	31.7526	31.7522	35.7726	31.0050	22.4638
$\underline{e}^+_{MEB.10}(\underline{Y})$		31.7530	31.7526	31.7522	35.7726	31.0050	22.4638
$\underline{e}_{PTE.05}(\underline{Y})$		31.7548	31.7548	31.7548	40.5139	32.1927	21.9560
$\underline{e}_{MEB.05}(\underline{Y})$		31.7530	31.7526	31.7522	39.6051	31.6586	21.6440
$\underline{e}^+_{MEB.05}(\underline{Y})$		31.7530	31.7526	31.7522	39.6051	31.6586	21.6440
$\underline{e}_{PTE.01}(\underline{Y})$		31.7548	31.7548	31.7548	47.0679	32.2514	20.6613
$\underline{e}_{MEB.01}(\underline{Y})$		31.7530	31.7526	31.7522	46.7918	32.1120	20.6191
$\underline{e}^+_{MEB.01}(\underline{Y})$		31.7530	31.7526	31.7522	46.7918	32.1120	20.6191

* Note: For $\underline{\nu} = (-4\,6\,8\,3\,5\,2\,7\,4)$, during the 1,000 replications, the null hypothesis, $\underline{\beta}_2 = \underline{0}$, was rejected in every replication for all α =.10, .05 and .01. Hence, $\underline{e}_{PTE\alpha}(\underline{Y}) = LSE$, $\underline{e}_{MEB\alpha}(\underline{Y}) = \underline{e}_{EB}(\underline{Y})$ and $\underline{e}^+_{MEB\alpha}(\underline{Y}) = \underline{e}^+_{EB}(\underline{Y})$.

Table 2.5: Six-week heights of chrysanthemum plants in pots treated with two growth retardant chemicals at six rates of application.

Block 1

	Rate					
Chemical	5	10	20	40	80	160
1	55.0	47.0	52.5	46.5	47.0	37.0
	57.5	54.0	51.0	44.5	41.0	48.0
2	54.0	47.5	38.0	44.0	25.0	4.50
	51.0	53.5	44.5	26.5	23.5	21.5

Block 5

	Rate					
Chemical	5	10	20	40	80	160
1	57.5	50.5	50.0	47.5	35.5	20.0
	48.0	55.0	53.0	32.5	40.0	32.0
2	45.0	52.5	32.0	43.0	31.0	26.0
	56.0	50.0	45.0	37.0	28.0	22.0

Table 2.5. The continuous nature of the variable rate is not incorporated in the results to be reported in this paper, because we use the example to illustrate behavior of the estimators rather than to illustrate how the data should be analyzed for publication in the horticultural literature.

The model assumed for a given block is

$$y_{ijk} = \mu + \alpha_i + \gamma_j + \delta_{ij} + \epsilon_{ijk}, \tag{2.92}$$

where y_{ijk} is the six week height of the kth plant treated with rate j of chemical i. We parameterize the model with $\Sigma_i \alpha_i = \Sigma_j \gamma_j = \Sigma_i \delta_{ij} = \Sigma_j \delta_{ij} = 0$, thereby eliminating any parameter with subscript $i = 2$ or $j = 6$. Thus, $\underline{\beta}_1^T = (\mu, \alpha_1, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5)$, $\underline{\beta}_2^T = (\delta_{11}, \delta_{12}, \delta_{13}, \delta_{14}, \delta_{15})$ and $\underline{\beta}^T = (\underline{\beta}_1^T, \underline{\beta}_2^T)$.

In matrix notation, the model is

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{141} \\ y_{142} \\ y_{151} \\ y_{152} \\ y_{161} \\ y_{162} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{231} \\ y_{232} \\ y_{241} \\ y_{242} \\ y_{251} \\ y_{252} \\ y_{261} \\ y_{262} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 \\ 1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \gamma_5 \\ \delta_{11} \\ \delta_{12} \\ \delta_{13} \\ \delta_{14} \\ \delta_{15} \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{131} \\ \epsilon_{132} \\ \epsilon_{141} \\ \epsilon_{142} \\ \epsilon_{151} \\ \epsilon_{152} \\ \epsilon_{161} \\ \epsilon_{162} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \\ \epsilon_{231} \\ \epsilon_{232} \\ \epsilon_{241} \\ \epsilon_{242} \\ \epsilon_{251} \\ \epsilon_{252} \\ \epsilon_{261} \\ \epsilon_{262} \end{bmatrix}. \tag{2.93}$$

The null hypothesis $H_0 : \underline{\beta}_2 = \underline{0}$ specifies no interaction between chemical and rate. Table 2.6 contains F statistics and observed significance probabilities for $H_0 : \underline{\beta}_2 = \underline{0}$ for each block. Also included in Table 2.6 are the sums of squared residuals from the LSE for the full model (SSE_{LSE}), the LSE for the reduced model (SSE_{RLSE}), the EB estimator (SSE_{EB}) and the hierarchical Bayes estimator with $f_1 = 2$ (SSE_{HB}). The Euclidean distances between the EB or the HB estimators from the least squares estimators for the full as well as the reduced models are also provided. These distances are noted as $D_{LSE,EB}$, $D_{RLSE,EB}$, $D_{LSE,HB}$ and $D_{RLSE,HB}$, ($D_{LSE,EB} = \| \underline{X}\tilde{\underline{\beta}} - \underline{e}_{EB}(\underline{Y}) \|^2$, *etc.*). The test statistic for interaction ($F = 3.10$, $p = 0.0503$) in block one is essentially significant at the .05 level ($F_{.05} = 3.11$), but the test statistic for interaction ($F = .70$, $p = .6318$) in block five is

Table 2.6: Sum of squared errors and the Euclidean distances for different estimators.

F for $H_0 : \underline{\beta}_2 = \underline{0}$	Block 1 $3.10(p = 0.0530)$	Block 5 $0.70(p = 0.6318)$
SSE_{LSE}	451.6	433.0
SSE_{RLSE}	1034.8	559.9
SSE_{EB}	467.7	500.9
SSE_{HB}	471.7	458.2
$D_{LSE,EB}$	16.10	67.90
$D_{RLSE,EB}$	405.7	9.20
$D_{LSE,HB}$	20.10	25.20
$D_{RLSE,HB}$	386.7	39.0

not significant at any meaningful level. Correspondingly, in block one the sum of squared residuals for the reduced model (1034.8) is much larger than the sum of squared residuals for the full model (451.6), and in block five the sum of squares for the reduced model (559.9) is only nominally larger than the sum of squared residuals for the full model (433.0). But in both blocks, the sum of squared residuals for the EB estimator (467.7 in block one and 500.9 in block five) is not much larger than the sum of squared residuals for the complete model. For the HB estimator with $f_1 = 2$, the sum of squared residuals (471.7 in block one and 458.2 in block five) behaves similarly as the EB estimator.

The Euclidean distances in block one ($D_{LSE,EB} = 16.1$, $D_{RLSE,EB} = 405.7$, $D_{LSE,HB} = 20.1$ and $D_{RLSE,HB} = 386.7$) shows that the distances between the EB or the HB estimators from the LSE for the full model is much less than the distances between the the EB or the HB estimators from the LSE for the reduced model while interaction is present.

But in block five the Euclidean distances ($D_{LSE,EB} = 67.9$, $D_{RLSE,EB} = 9.2$, $D_{LSE,HB} = 25.2$ and $D_{RLSE,HB} = 39.0$) between the EB estimator from the LSE

for the reduced model is less than the distance between the EB estimator from the LSE for the full model while the interaction is insignificant. But surprisingly, the HB estimator is still closer to the full model LSE than the reduced model LSE.

Predictive values for the full model ($\underline{X}\tilde{\beta}$), for the reduced model ($\underline{X}_1\hat{\beta}_1$), for the EB estimator $\underline{e}_{EB}(\underline{Y})$ and for the HB estimator $\underline{e}_{HB}(\underline{Y})$ are presented in Table 2.7 and plotted in Figure 2.1. In block one, wherein interaction is present, the EB and the HB estimates differ relatively little from the LSE's for the full model. But in block five, wherein interaction is insignificant, the EB estimates are close to the LSE's for the reduced model.

Results for the preliminary test estimators, $\underline{e}_{PTE.10}(\underline{Y})$, $\underline{e}_{PTE.05}(\underline{Y})$, and $\underline{e}_{PTE.01}(\underline{Y})$ are not shown explicitly in Table 2.7 because their values are equal to either the LSE $\underline{X}\tilde{\beta}$ or the $RLSE$ $\underline{X}_1\hat{\beta}_1$, depending on the level of significance of the F test for interaction. In block five the F test was not significnat at any meaningful level; so all the preliminary test estimators are equal to the LSE for the reduced model. In block one, the test for interaction ($F = 3.10$) fell short of significance at the .05 level ($F_{.05} = 3.11$) by a very narrow margin. Thus $\underline{e}_{PTE.05}(\underline{Y})$ and $\underline{e}_{PTE.01}(\underline{Y})$ are both equal to the $RLSE$ in block one, but $\underline{e}_{PTE.10}(\underline{Y})$ is equal to LSE for the full model. This illustrates a fallacy in the PT estimators, especially when the α is set at a relatively low value such as .05 or .01. Interaction between chemical and rate is evidently present in block one, although the F test very slightly fails to detect it at the .05 (or smaller) level. This leads to values of $\underline{e}_{PTE.05}(\underline{Y})$ and $\underline{e}_{PTE.01}(\underline{Y})$ that are inferior to the LSE because of the apparent bias. In block one, $\underline{e}_{PTE.05}(\underline{Y})$ has values that are very different from values that would be obtained if F were only slightly larger. The EB or HB estimators do not suffer from this drawback.

Table 2.7: Predicted values for different estimators.

Block 1

Chem	Rate	$\underline{X}\tilde{\underline{\beta}}$	$\underline{X}_1\hat{\underline{\beta}}_1$	$\underline{e}_{EB}(\underline{Y})$	$\underline{e}_{HB}(\underline{Y})$
1	5	56.25	60.25	56.96	57.04
1	10	50.50	56.65	51.52	51.64
1	20	51.75	52.65	51.90	51.92
1	40	45.50	46.52	45.67	45.69
1	80	44.00	40.27	43.38	43.31
1	160	42.50	33.90	41.07	40.90
2	5	52.50	48.23	51.79	51.71
2	10	50.50	44.35	49.48	49.36
2	20	41.25	40.35	41.10	41.08
2	40	35.25	34.23	35.08	35.06
2	80	24.25	27.98	24.87	24.94
2	160	13.00	21.60	14.43	14.60

Block 5

Chem	Rate	$\underline{X}\tilde{\underline{\beta}}$	$\underline{X}_1\hat{\underline{\beta}}_1$	$\underline{e}_{EB}(\underline{Y})$	$\underline{e}_{HB}(\underline{Y})$
1	5	52.75	53.88	53.57	53.25
1	10	52.75	54.25	53.85	53.42
1	20	51.50	47.25	43.89	49.16
1	40	40.00	42.25	41.65	41.00
1	80	37.75	35.88	36.38	36.91
1	160	26.00	27.25	26.91	26.56
2	5	50.50	49.38	49.68	50.00
2	10	51.25	49.75	50.15	50.58
2	20	38.50	42.75	41.61	40.39
2	40	40.00	37.75	38.35	39.00
2	80	29.50	31.38	30.87	30.36
2	160	24.00	22.75	23.09	23.44

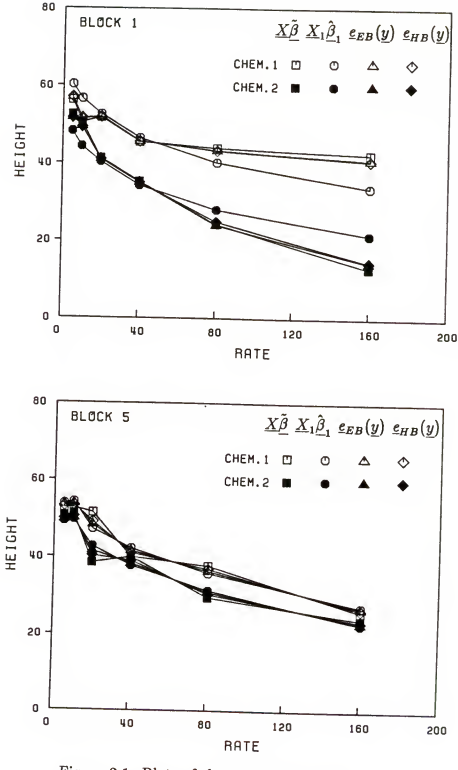

Figure 2.1: Plots of the predictive values;

block 1: interaction, and block 5: no interaction.

CHAPTER 3
EMPIRICAL BAYES ESTIMATION OF MULTIVARIATE REGRESSION MODEL I

3.1 Introduction

The linear model for m distinct individuals is often useful in growth and response curve studies. If it is suspected that there are no interaction effects involved, then, there may be a choice between the following two models:

$$\underline{Y}_k = \underline{X}_{k1}\underline{\beta}_{k1} + \underline{X}_{k2}\underline{\beta}_{k2} + \underline{e}_k, \tag{3.1}$$

$$\underline{Y}_k = \underline{X}_{k1}\underline{\beta}_{k1} + \underline{e}_k, \tag{3.2}$$

where $\underline{Y}_k$ is a $n_k \times 1$ vector of responses of the kth individual over the n_k different occasions, $\underline{X}_i$'s are the $n_k \times p_i$ design matrices of full rank $p_i < n_k$ for $i = 1, 2$. $\underline{\beta}_{k1}$ is the $p_1 \times 1$ vector of main effects coefficients, and $\underline{\beta}_{k2}$ is the $p_2 \times 1$ vector of interaction coefficients for the kth individual. The vectors $\underline{e}_k$'s are independently distributed as $N(\underline{0}, \sigma^2\underline{V}_k)$, where $\underline{V}_k$'s are the known p.d. matrices. However, the above the analysis follows trivially from Chapter 2. If in addition the main effects are suspected to be all equal, then, there may be a choice between the model (3.1) and the following one:

$$\underline{Y}_k = \underline{X}_{k1}\underline{\beta}_{1\cdot} + \underline{e}_k. \tag{3.3}$$

To determine the choice between the model (3.1) and the model (3.3), a classical way is testing $H_0 : \underline{\beta}_{k1} = \underline{\beta}_{1\cdot}$ and $\underline{\beta}_{k2} = \underline{0}$ for all k. The model (3.1) is utilized if the null hypothesis is rejected at a desired level of significance (for example, $\alpha = .05$), and the model (3.3) is chosen when the opposite is true. However, the above procedure suffers from the drawback that there is no general way of

incorporating the degree of evidence for or against the null hypothesis in order to arrive at the estimators.

For the problem of best fitting of a given set of data by the least squares methods, the model (3.1) is always superior to the model (3.3) i.e. $\sum_{k=1}^{m} \| \underline{Y}_k - \underline{X}_k\tilde{\underline{\beta}}_k \|^2 \leq \sum_{k=1}^{m} \| \underline{Y}_k - \underline{X}_{k1}\hat{\underline{\beta}}_{1.} \|^2$. Note that $\tilde{\underline{\beta}}_k = (\underline{X}_k^T\underline{V}_k^{-1}\underline{X}_k)^{-1}\underline{X}_k\underline{V}_k^{-1}\underline{Y}_k$ denotes the LSE of $\underline{\beta}_k$ based on the model (3.1) and $\hat{\underline{\beta}}_{1.} = (\sum_{k=1}^{m} \underline{C}_{k11})^{-1} \sum_{k=1}^{m} \underline{C}_{k11}\hat{\underline{\beta}}_{k1}$ with $\underline{C}_{k11} = \underline{X}_{k1}^T\underline{V}_k^{-1}\underline{X}_{k1}$ and $\hat{\underline{\beta}}_{k1} = \underline{C}_{k11}^{-1}\underline{X}_{k1}^T\underline{V}_k^{-1}\underline{Y}_k$. However, when estimators of $(\underline{X\beta})^* = ((\underline{X}_1\underline{\beta}_1)^T, \cdots, (\underline{X}_m\underline{\beta}_m)^T)^T$ are compared on the basis of general quadratic loss,

$$L(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{a}_1, \cdots, \underline{a}_m) = \sum_{k=1}^{m} (\underline{a}_k - \underline{X}_k\underline{\beta}_k)^T\underline{Q}_k(\underline{a}_k - \underline{X}_k\underline{\beta}_k), \qquad (3.4)$$

where $\underline{Q}_k$'s are some known p.d. matrix, the LSE of $(\underline{X\beta})^*$ based on the model (3.1) may lose their superiority.

In this chapter, the empirical Bayes (EB) estimators of $(\underline{X\beta})^*$ are proposed by some given appropriate g-priors. The EB estimatiors serve as a weighted average of the LSE of $\underline{X}_k\underline{\beta}_k$ $k = 1, \cdots, m$, based on the model (3.1) and the model (3.3). This is of practical benefit especially when the weights are so chosen such that the compromise estimator of $\underline{X}_k\underline{\beta}_k$ leans more towards $\underline{X}_k\tilde{\underline{\beta}}_k$ for all k if the observed F ratio is large, and to $\underline{X}_{k1}\hat{\underline{\beta}}_{1.}$ if the observed F ratio is small. In Chapter 4, a choice between the model (3.1) and another reduced multivariate model will be discussed, also, for reaching a proper empirical Bayes (EB) estimator, a different g-prior; also based on the prior derived by Reinsel (1985), will be discussed.

A popular method of achieving a compromise based on the full and the reduced models is the use of the preliminary test estimators (PTE) based on the rejection or acceptance of the null hypothesis H_0. It shown in Section 3.2 that for every PTE of $\underline{X}_k\underline{\beta}_k$'s, there is a corresponding modified EB estimator which dominates the PTE. This is not a surprising consequence, since the PTE also does not take into account the degree of evidence for or against the null hypothesis $H_0 : \underline{\beta}_{k1} = \underline{\beta}_{1.}$ and $\underline{\beta}_{k2} = \underline{0}$ for all k.

Section 3.2 introduces the EB estimators and their positive part versions. Several theorems are given which describe the frequentist as well as the Bayesian properities of the proposed estimators. It is shown that there are EB estimators of $(\underline{X\beta})^*$ which always dominate $\underline{e}^*_{LSE} = ((\underline{X}_1\tilde{\underline{\beta}}_1)^T, \cdot, (\underline{X}_m\tilde{\underline{\beta}}_m)^T)^T$ under squared error loss. Section 3.3 introduces several estimators of $\underline{\beta}_{k1}$. These estimators are proposed such that they shrink the unrestricted LSE $\tilde{\underline{\beta}}_{k1}$ to the restricted LSE $\hat{\underline{\beta}}_{k1}$ or $\hat{\underline{\beta}}_{.1}$, with different shrinkage coefficients for some k. Under squared error loss, their frequentist risks and Bayes risks are compared. Finally, in Section 3.4, in a multivariate regression example, the different estimators of $(\underline{X\beta})^*$ are compared according to their simulated frequentist and Bayes risks.

3.2 The Empirical Bayes Approcah

In order to develop an EB procedure, a Bayes estimator of $\underline{X}_k\underline{\beta}_k$ for $k = 1, \cdots, m$, should first be obtained. Based on the null hopothesis $H_0 : \underline{\beta}_{k1} = \underline{\beta}_{.1}$ and $\underline{\beta}_{k2} = \underline{0}$ for all k, a g-prior distribution of $\underline{\beta}_k$'s for the general mulitivariate regression model in (3.1) is

$$\underline{\beta}_k \sim^{ind.} N_p(\underline{\nu}, \tau^2(\underline{X}_k^T\underline{V}_k^{-1}\underline{X}_k)^{-1}), \tag{3.5}$$

where $\underline{\nu} = \begin{pmatrix} \underline{\nu}_1 \\ \underline{0} \end{pmatrix}$ with the dimensions of $\underline{\nu}_1$ and $\underline{0}$ equal to $p_1 \times 1$, $p_2 \times 1$, respectively, and $p_1 + p_2 = p$.

With the above g-prior, the posterior of $\underline{\beta}_k$ given $\underline{Y}_k$ is

$$N_p(\underline{\nu} + (1-B)(\tilde{\underline{\beta}}_k - \underline{\nu}), \sigma^2(1-B)(\underline{X}_k^T\underline{V}_k^{-1}\underline{X}_k)^{-1}), k = 1, \cdots, m, \tag{3.6}$$

where $\tilde{\underline{\beta}}_k = (\underline{X}_k^T\underline{V}_k^{-1}\underline{X}_k)^{-1}\underline{X}_k^T\underline{V}_k^{-1}\underline{Y}_k$ is the $BLUE$ of $\underline{\beta}_k$, and $B = \sigma^2/(\sigma^2 + \tau^2)$. By expressing $\tilde{\underline{\beta}}_k^T = (\tilde{\underline{\beta}}_{k1}^T, \tilde{\underline{\beta}}_{k2}^T)$, the Bayes estimator of $\underline{X}_k\underline{\beta}_k$ under any quadratic loss (3.4) is

$$\begin{aligned} & \underline{e}_{kB}(\underline{Y}_k) = E(\underline{X}_k\underline{\beta}_k|\underline{Y}_k) \\ = \; & \underline{X}_{k1}[\,\underline{\nu}_1 + (1-B)(\tilde{\underline{\beta}}_{k1} - \underline{\nu}_1)\,] + (1-B)\underline{X}_{k2}\tilde{\underline{\beta}}_{k2}, \text{for } k = 1, \cdots, m. \end{aligned} \tag{3.7}$$

In an empirical Bayes framework, the parameters $\underline{\nu}_1$, σ^2 and τ^2 are unknown, and need to be estimated from the marginal distribution of $\underline{Y}_1, \cdots, \underline{Y}_m$. Note that marginally

$$\underline{Y}_k \sim^{ind.} N_{nk}(\underline{X}_{k1}\underline{\nu}_1, \sigma^2\underline{V}_k + \tau^2\underline{P}_{\underline{X}_k}), \tag{3.8}$$

where $\underline{P}_{\underline{X}_k} = \underline{X}_k(\underline{X}_k^T\underline{V}_k^{-1}\underline{X}_k)^{-1}\underline{X}_k^T$. By using Exercise 2.9, p.33 of Rao (1973), and recalling $B = \sigma^2/(\sigma^2 + \tau^2)$, it can be shown that

$$\begin{aligned}&(\sigma^2\underline{V}_k + \tau^2\underline{P}_{\underline{X}_k})^{-1}\\ &= \sigma^{-2}[\underline{V}_k^{-1} - (1-B)\underline{V}_k^{-1}\underline{P}_{\underline{X}_k}\underline{V}_k^{-1}].\end{aligned} \tag{3.9}$$

Now, let $\underline{C}_{kij} = \underline{X}_{ki}^T\underline{V}_k^{-1}\underline{X}_{kj}$ $\underline{C}_{ki\cdot j} = \underline{C}_{kii} - \underline{C}_{ki\underline{j}}\underline{C}_{kjj}^{-1}\underline{C}_{kji}$ $(i,j = 1,2$ and $k = 1,\cdots,m)$. Under the model (3.3), the $BLUE$ of $\underline{\beta}_{1\cdot}$ is $\hat{\underline{\beta}}_{1\cdot} = (\sum_{k=1}^m \underline{C}_{k11})^{-1}\sum_{k=1}^m \underline{C}_{k11}\hat{\underline{\beta}}_{k1}$, with $\hat{\underline{\beta}}_{k1} = (\underline{X}_{k1}^T\underline{V}_k^{-1}\underline{X}_{k1})^{-1}\ \underline{X}_{k1}^T\underline{V}_k^{-1}\underline{Y}_k$. Then, based on the normal equation

$$\underline{C}_{k11}\hat{\underline{\beta}}_{k1} = \underline{C}_{k11}\tilde{\underline{\beta}}_{k1} + \underline{C}_{k12}\tilde{\underline{\beta}}_{k2}, \tag{3.10}$$

the following identity can be obtained

$$\tilde{\underline{\beta}}_{k1} = \hat{\underline{\beta}}_{k1} - \underline{C}_{k11}^{-1}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2}. \tag{3.11}$$

Since $\underline{P}_{\underline{X}_k}\underline{V}_k^{-1}\underline{X}_{ki} = \underline{X}_{ki}$ $(i = 1,2)$, using the idempotency of $\underline{V}_k^{-1}\underline{P}_{\underline{X}_k}$, and (3.9), one gets

$$\begin{aligned}&(\underline{Y}_k - \underline{X}_{k1}\underline{\nu}_1)^T(\sigma^2\underline{V}_k + \tau^2\underline{P}_{\underline{X}_k})^{-1}(\underline{Y}_k - \underline{X}_{k1}\underline{\nu}_1)\\ &= \sigma^{-2}\{(\underline{Y}_k - \underline{X}_k\tilde{\underline{\beta}}_k + \underline{X}_k\begin{pmatrix}\tilde{\underline{\beta}}_{k1} - \underline{\nu}_1\\ \underline{\beta}_{k2}\end{pmatrix})^T\\ &\quad (\underline{V}_k^{-1} - (1-B)\underline{V}_k^{-1}\underline{P}_{\underline{X}_k}\underline{V}_k^{-1})\\ &\quad (\underline{Y}_k - \underline{X}_k\tilde{\underline{\beta}}_k + \underline{X}_k\begin{pmatrix}\tilde{\underline{\beta}}_{k1} - \underline{\nu}_1\\ \underline{\beta}_{k2}\end{pmatrix})\}\\ &= \sigma^{-2}\{(\underline{Y}_k - \underline{X}_k\tilde{\underline{\beta}}_k)^T\underline{V}_k^{-1}(\underline{Y}_k - \underline{X}_k\tilde{\underline{\beta}}_k)\\ &+ B((\tilde{\underline{\beta}}_{k1} - \underline{\nu}_1)^T\ \tilde{\underline{\beta}}_{k2}^T)(\underline{X}_k^T\underline{V}_k^{-1}\underline{X}_k)\begin{pmatrix}\tilde{\underline{\beta}}_{k1} - \underline{\nu}_1\\ \underline{\beta}_{k2}\end{pmatrix}\}.\end{aligned}$$

Thus, adding all k components of the above equation, and using the identity (3.11), one gets

$$\begin{aligned}
&\sum_{k=1}^{m}(\underline{Y}_k - \underline{X}_{k1}\underline{\nu}_1)^T(\sigma^2\underline{V}_k + \tau^2\underline{P}_{\underline{X}_k})^{-1}(\underline{Y}_k - \underline{X}_{k1}\underline{\nu}_1) \\
&= \sigma^{-2}\{SSE + B\sum_{k=1}^{m}[\,(\underline{\hat{\beta}}_{k1} - \underline{\nu}_1)^T \;\; \underline{\tilde{\beta}}_{k2}^T\,]\begin{bmatrix} \underline{C}_{k11} & \underline{C}_{k12} \\ \underline{C}_{k21} & \underline{C}_{k22} \end{bmatrix}\begin{bmatrix} \underline{\hat{\beta}}_{k1} - \underline{\nu}_1 \\ \underline{\beta}_{k2} \end{bmatrix}\}, \\
&= \sigma^{-2}\{SSE + B\{\sum_{k=1}^{m}(\underline{\hat{\beta}}_{k1} - \underline{\nu}_1)^T\underline{C}_{k11}(\underline{\hat{\beta}}_{k1} - \underline{\nu}_1) \\
&+ \sum_{k=1}^{m}\underline{\tilde{\beta}}_{k2}^T\underline{C}_{k22.1}\underline{\tilde{\beta}}_{k2}\}\}, \qquad (3.12)
\end{aligned}$$

where

$$\begin{aligned}
&\sum_{k=1}^{m}(\underline{\hat{\beta}}_{k1} - \underline{\nu}_1)^T\underline{C}_{k11}(\underline{\hat{\beta}}_{k1} - \underline{\nu}_1) \\
&= (\underline{\nu}_1 - \underline{\hat{\beta}}_{1.})^T(\sum_{k=1}^{m}\underline{C}_{k11})(\underline{\nu}_1 - \underline{\hat{\beta}}_{1.}) \\
&+ \sum_{k=1}^{m}(\underline{\hat{\beta}}_{k1} - \underline{\hat{\beta}}_{1.})^T\underline{C}_{k11}(\underline{\hat{\beta}}_{k1} - \underline{\hat{\beta}}_{1.}), \qquad (3.13)
\end{aligned}$$

and $SSE = \sum_{k=1}^{m}(\underline{Y}_k - \underline{X}_k\underline{\tilde{\beta}}_k)^T\underline{V}_k^{-1}(\underline{Y}_k - \underline{X}_k\underline{\tilde{\beta}}_k)$, the usual error sum of squares. Let $SSH_1 = \sum_{k=1}^{m}(\underline{\hat{\beta}}_{k1} - \underline{\hat{\beta}}_{1.})^T\ \underline{C}_{k11}(\underline{\hat{\beta}}_{k1} - \underline{\hat{\beta}}_{1.})$, and $SSH_2 = \sum_{k=1}^{m}\underline{\tilde{\beta}}_{k2}^T\underline{C}_{k22.1}\underline{\tilde{\beta}}_{k2}$. The summation of SSH_1 and SSH_2 is the sum of squares due to the hypothesis $H_0 : \underline{\beta}_{k1} = \underline{\beta}_{1.}$ and $\underline{\beta}_{k2} = \underline{0}$ for all k. From (3.12) and (3.13), it is easy to see that $(\underline{\hat{\beta}}_{1.}, SSH_1 + SSH_2, SSE)$ is complete sufficient for $\underline{\nu}_1, \tau^2, \sigma^2$.

Define $\underline{C} = \sum_{k=1}^{m}\underline{C}_{k11}$. Since

$$\begin{aligned}
SSH_1 &= (\underline{\hat{\beta}}_{11}^T, \cdots, \underline{\hat{\beta}}_{m1}^T) \\
&\times \Bigg(\begin{bmatrix} \underline{C}_{111} & \cdots & \underline{0} \\ \vdots & \ddots & \vdots \\ \underline{0} & \cdots & \underline{C}_{m11} \end{bmatrix} - \begin{bmatrix} \underline{C}_{111}\underline{C}^{-1}\underline{C}_{111} & \cdots & \underline{C}_{111}\underline{C}^{-1}\underline{C}_{m11} \\ \vdots & \ddots & \vdots \\ \underline{C}_{m11}\underline{C}^{-1}\underline{C}_{111} & \cdots & \underline{C}_{m11}\underline{C}^{-1}\underline{C}_{m11} \end{bmatrix}\Bigg) \\
&\times \begin{bmatrix} \underline{\hat{\beta}}_{11} \\ \vdots \\ \underline{\hat{\beta}}_{m1} \end{bmatrix},
\end{aligned}$$

$$V\left(\begin{bmatrix} \hat{\underline{\beta}}_{11} \\ \vdots \\ \hat{\underline{\beta}}_{m1} \end{bmatrix}\right) = (\sigma^2+\tau^2)\begin{bmatrix} \underline{C}_{111}^{-1} & \cdots & \underline{0} \\ \vdots & \ddots & \vdots \\ \underline{0} & \cdots & \underline{C}_{m11}^{-1} \end{bmatrix},$$

and

$$\left(\begin{bmatrix} \underline{C}_{111} & \cdots & \underline{0} \\ \vdots & \ddots & \vdots \\ \underline{0} & \cdots & \underline{C}_{m11} \end{bmatrix} - \begin{bmatrix} \underline{C}_{111}\underline{C}^{-1}\underline{C}_{111} & \cdots & \underline{C}_{111}\underline{C}^{-1}\underline{C}_{m11} \\ \vdots & \ddots & \vdots \\ \underline{C}_{m11}\underline{C}^{-1}\underline{C}_{111} & \cdots & \underline{C}_{m11}\underline{C}^{-1}\underline{C}_{m11} \end{bmatrix}\right)$$
$$\times \begin{bmatrix} \underline{C}_{111}^{-1} & \cdots & \underline{0} \\ \vdots & \ddots & \vdots \\ \underline{0} & \cdots & \underline{C}_{m11}^{-1} \end{bmatrix}$$
$$= \begin{bmatrix} \underline{I}_{p_1} & \cdots & \underline{0} \\ \vdots & \ddots & \vdots \\ \underline{0} & \cdots & \underline{I}_{p_1} \end{bmatrix} - \begin{bmatrix} \underline{C}_{111}\underline{C}^{-1} & \cdots & \underline{C}_{111}\underline{C}^{-1} \\ \vdots & \ddots & \vdots \\ \underline{C}_{m11}\underline{C}^{-1} & \cdots & \underline{C}_{m11}\underline{C}^{-1} \end{bmatrix}. \tag{3.14}$$

The right hand side of (3.14) is an idempotent matrix with trace equal to $(m-1)p_1$. Thus, $SSH_1 \sim (\tau^2+\sigma^2)\chi^2_{(m-1)p_1}$. Since $\hat{\underline{\beta}}_{k1}$'s are indepedent of $\hat{\underline{\beta}}_{k2}$'s, (see Lemma 2 in Ghosh, Saleh, Sen, 1987), and $SSH_2 \sim (\tau^2+\sigma^2)\chi^2_{mp_2}$, thus, $SSH_1 + SSH_2 \sim (\tau^2+\sigma^2)\chi^2_{mp-p_1}$. The $UMVUE$ of $\sigma^{-2}B = (\tau^2+\sigma^2)^{-1}$ is then given by $(mp-p_1-2)/(SSH_1+SSH_2)$ for $(mp-p_1) \geq 3$ and $p = p_1+p_2$. The best scale invariant estimate of σ^2 is $SSE/(N-mp+2)$, where $N = \sum_{k=1}^m n_k$, and the $UMVUE$ of $\underline{\nu}_1$ is $\hat{\underline{\beta}}_{1.}$. Substitutes of all these estimators for the unknown parameters in (3.7) gives

$$\begin{aligned} &\underline{e}_{kEB}(\underline{Y}_k) \\ &= \underline{X}_{k1}(\hat{\underline{\beta}}_{1.} + (1 - \frac{mp-p_1-2}{F})(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})) \\ &+ (1 - \frac{mp-p_1-2}{F})\underline{X}_{k2}\tilde{\underline{\beta}}_{k2},\ k = 1,\cdots,m, \end{aligned} \tag{3.15}$$

where $F = (N-mp+2)(SSH_1+SSH_2)/SSE$, a constant multiple of the F ratio for testing $H_0 : \underline{\beta}_{k1} = \underline{\beta}_{1.}$ and $\underline{\beta}_{k2} = \underline{0}$ for all k.

In parctice, it is much more appropriate to use the positive part EB estimator, $\underline{e}^{+*}_{EB} = (\underline{e}^{+T}_{1EB}, \cdots, \underline{e}^{+T}_{mEB})^T$, because B is estimated by the quantity $(mp-p_1-2)/F$ which can be greater than 1.

$$
\begin{aligned}
& \underline{\varepsilon}^{+}_{kEB}(\underline{Y}_k) \\
= \; & \underline{X}_{k1}(\hat{\underline{\beta}}_{1.} + (1 - \frac{mp - p_1 - 2}{F})^{+}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})) \\
+ \; & (1 - \frac{mp - p_1 - 2}{F})^{+}\underline{X}_{k2}\tilde{\underline{\beta}}_{k2}, \; k = 1, \cdots, m
\end{aligned} \tag{3.16}
$$

where $a^{+} = \max(a, 0)$.

Note from (3.16) that for very large F values signifying substantial departure from $H_0 : \underline{\beta}_{k1} = \underline{\beta}_{1.}$ and $\underline{\beta}_{k2} = \underline{0}$ for all k, $\underline{\varepsilon}^{+}_{kEB}$ is very close to $\underline{X}_k\tilde{\underline{\beta}}_k$, whereas for very small F values signifying enough support for H_0, $\underline{\varepsilon}^{+}_{kEB}$ is very close to $\underline{X}_{k1}\hat{\underline{\beta}}_{1.}$ for all k. When there is no clear-cut decision for or against H_0, $\underline{\varepsilon}^{+}_{kEB}$ is in some sense a weighted average of $\underline{X}_{k1}\hat{\underline{\beta}}_{1.}$ and $\underline{X}_k\tilde{\underline{\beta}}_k$, with the weights being adaptively determined by the data.

Next, the estimators $\underline{\varepsilon}^{*}_{EB} = (\underline{\varepsilon}^{T}_{1EB}, \cdots, \underline{\varepsilon}^{T}_{mEB})^{T}$ and $\underline{\varepsilon}^{++}_{EB}$ will be evaluated according to the criteria of frequentist and Bayes risks. First the frequentist criterion will be utilized.

The general class of estimators of $\underline{X}_k\underline{\beta}_k$ is

$$
\begin{aligned}
& \underline{\varepsilon}_{k\phi}(\underline{Y}_k) \\
= \; & \underline{X}_{k1}(\hat{\underline{\beta}}_{1.} + (1 - \frac{\phi(F)}{F})(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})) + (1 - \frac{\phi(F)}{F})\underline{X}_{k2}\tilde{\underline{\beta}}_{k2}, \\
& k = 1, \cdots, m.
\end{aligned} \tag{3.17}
$$

The following theorem provides the risk expression for $\underline{\varepsilon}_{k\phi}$ under the loss (3.4) with $\underline{Q}_k = \underline{V}_k^{-1}$. The risk is denoted by $R(\underline{X}_1\underline{\beta}_k, \cdots, \underline{X}_m\underline{\beta}_m ; \underline{\varepsilon}_{1\phi}, \cdots, \underline{\varepsilon}_{m\phi})$.

Theorem 3.2.1. If ϕ is a differentiable function, then, under the loss (3.4) with $\underline{Q}_k = \underline{V}_k^{-1}$,

$$
\begin{aligned}
& R(\underline{X}_1\underline{\beta}_k, \cdots, \underline{X}_m\underline{\beta}_m; \underline{\varepsilon}_{1\phi}, \cdots, \underline{\varepsilon}_{m\phi}) \\
= \; & R(\underline{X}_1\underline{\beta}_k, \cdots, \underline{X}_m\underline{\beta}_m; \underline{X}_1\tilde{\underline{\beta}}_1, \cdots, \underline{X}_m\tilde{\underline{\beta}}_m)
\end{aligned}
$$

$$
\begin{aligned}
&- \ 2\sigma^2 E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{\frac{\phi(F)}{F}(mp-p_1)+2(\frac{\phi'(F)}{F}-\frac{\phi(F)}{F^2})F\} \\
&+ \ E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{\frac{\phi^2(F)}{F^2}(SSH_1+SSH_2)\}. \qquad (3.18)
\end{aligned}
$$

Proof of Theorem 3.2.1. By using (3.11), it can be shown that

$$
\begin{aligned}
&\underline{e}_{k\phi}-\underline{X}_k\underline{\beta}_k \\
&= \ \underline{X}_{k1}[\,\hat{\underline{\beta}}_{k1}-E(\hat{\underline{\beta}}_{k1})-\frac{\phi(F)}{F}(\hat{\underline{\beta}}_{k1}-\hat{\underline{\beta}}_{1.})\,] \\
&+ \ (\underline{X}_{k2}-\underline{X}_{k1}\underline{C}_{k11}^{-1}\underline{C}_{k12})[\,(1-\frac{\phi(F)}{F})\tilde{\underline{\beta}}_{k2}-\underline{\beta}_{k2}\,]. \qquad (3.19)
\end{aligned}
$$

Then,

$$
\begin{aligned}
&\sum_{k=1}^{m} E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}(\underline{e}_{k\phi}-\underline{X}_k\underline{\beta}_k)^T\underline{V}_k^{-1}(\underline{e}_{k\phi}-\underline{X}_k\underline{\beta}_k) \\
&= \ \sum_{k=1}^{m} E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{[\,\hat{\underline{\beta}}_{k1}-E(\hat{\underline{\beta}}_{k1})-\frac{\phi(F)}{F}(\hat{\underline{\beta}}_{k1}-\hat{\underline{\beta}}_{1.})\,]^T \\
&\quad \underline{C}_{k11}[\,\hat{\underline{\beta}}_{k1}-E(\hat{\underline{\beta}}_{k1})-\frac{\phi(F)}{F}(\hat{\underline{\beta}}_{k1}-\hat{\underline{\beta}}_{1.})\,] \\
&+ \ (\tilde{\underline{\beta}}_{k2}-\underline{\beta}_{k2}-\frac{\phi(F)}{F}\tilde{\underline{\beta}}_{k2})^T\underline{C}_{k22.1}(\tilde{\underline{\beta}}_{k2}-\underline{\beta}_{k2}-\frac{\phi(F)}{F}\tilde{\underline{\beta}}_{k2})\} \\
&= \ mp\sigma^2-2E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{\frac{\phi(F)}{F}\sum_{k=1}^{m}[\,(\hat{\underline{\beta}}_{k1}-\hat{\underline{\beta}}_{1.})^T\underline{C}_{k11}(\hat{\underline{\beta}}_{k1}-E(\hat{\underline{\beta}}_{k1})) \\
&+ \ \tilde{\underline{\beta}}_{k2}^T\underline{C}_{k22.1}(\tilde{\underline{\beta}}_{k2}-\underline{\beta}_{k2})\,]\} \\
&+ \ E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{\frac{\phi^2(F)}{F^2}\sum_{k=1}^{m}[\,(\hat{\underline{\beta}}_{k1}-\hat{\underline{\beta}}_{1.})^T\underline{C}_{k11}(\hat{\underline{\beta}}_{k1}-\hat{\underline{\beta}}_{1.}) \\
&+ \ \tilde{\underline{\beta}}_{k2}^T\underline{C}_{k22.1}\tilde{\underline{\beta}}_{k2}\,]\} \qquad (3.20)
\end{aligned}
$$

A generalization of Stein's identity (Berger and Haff, 1983) gives

$$
\begin{aligned}
&E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{\frac{\phi(F)}{F}\sum_{k=1}^{m}[\,(\hat{\underline{\beta}}_{k1}-\hat{\underline{\beta}}_{1.})^T\underline{C}_{k11}(\hat{\underline{\beta}}_{k1}-E(\hat{\underline{\beta}}_{k1})) \\
&+ \ \tilde{\underline{\beta}}_{k2}^T\underline{C}_{k22.1}(\tilde{\underline{\beta}}_{k2}-\underline{\beta}_{k2})\,]\} \\
&= \ \sum_{k=1}^{m} E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{tr\,\{[\,\frac{\partial}{\partial\hat{\underline{\beta}}_{k1}}(\frac{\phi(F)}{F}(\hat{\underline{\beta}}_{k1}-\hat{\underline{\beta}}_{1.})^T)\,]\underline{C}_{k11}V(\hat{\underline{\beta}}_{k1})\}\} \\
&+ \ tr\,\{[\,\frac{\partial}{\partial\tilde{\underline{\beta}}_{k2}}(\frac{\phi(F)}{F}\tilde{\underline{\beta}}_{k2}^T)\,]\underline{C}_{k22.1}V(\tilde{\underline{\beta}}_{k2})\}\} \\
&= \ \sigma^2 E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{2(\frac{\phi'(F)}{F}-\frac{\phi(F)}{F^2})\frac{SSH_1+SSH_2}{SSE/(N-mp-2)}
\end{aligned}
$$

$$+ \quad \frac{\phi(F)}{F}((m-1)p_1 + mp_2)\} \tag{3.21}$$

Recall that $SSH_1 = \sum_{k=1}^{m}(\underline{\hat{\beta}}_{k1} - \underline{\hat{\beta}}_{\cdot 1})^T \underline{C}_{k11}(\underline{\hat{\beta}}_{k1} - \underline{\hat{\beta}}_{\cdot 1})$, $SSH_2 = \sum_{k=1}^{m} \underline{\tilde{\beta}}_{k2}^T \underline{C}_{k22.1}\underline{\tilde{\beta}}_{k2}$, and $F = ((SSH_1 + SSH_2)/SSE)(N - mp + 2)$. Since $R(\underline{X}_1\underline{\beta}_k, \cdots, \underline{X}_m\underline{\beta}_m; \underline{X}_1\underline{\tilde{\beta}}_1, \cdots, \underline{X}_m\underline{\tilde{\beta}}_m) = mp\sigma^2$, (3.18) can be obtained by combining (3.20) and (3.21). Theorem 3.2.1 is hence proved.

The following corollary to Theorem 3.2.1 provides sufficient conditions under which $\underline{e}_{\phi}^{*} = (\underline{e}_{1\phi}^T, \cdots, \underline{e}_{m\phi}^T)^T$ dominates $\underline{e}_{LSE}^{*}$.

Corollary 3.2.1. Suppose $\underline{Q}_k = \underline{V}_k^{-1}$ and $mp - p_1 \geq 3$; $\phi(F)$ is non-decreasing in F, and $0 < \phi(F) < 2(mp - p_1 - 2)$. Then,

$$R(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{e}_{1\phi}, \cdots, \underline{e}_{m\phi}) < R(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{X}_1\tilde{\underline{\beta}}_1, \cdots, \underline{X}_m\tilde{\underline{\beta}}_m).$$

Proof of Corollary 3.2.1. Let $h(F) = \phi(F)/F$. Applying (2.18) of Efron and Morris (1976) and (3.9) of Ghosh, Saleh, and Sen (1987), the last term of (3.18) can be expressed as

$$\begin{aligned} & E_{\underline{\beta}_1, \cdots, \underline{\beta}_m, \sigma^2}(h^2(F)F(SSE/(N - mp + 2))) \\ = \quad & \sigma^2 E_{\underline{\beta}_1, \cdots, \underline{\beta}_m, \sigma^2}\left(\frac{\phi^2(F)}{F} - \frac{4}{N - mp + 2}\phi(F)\phi'(F)\right). \end{aligned} \tag{3.22}$$

From (3.22), and using the property $\phi'(F) > 0$,

$$\begin{aligned} & R(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{e}_{1\phi}, \cdots, \underline{e}_{m\phi}) \\ \leq \quad & R(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{X}_1\tilde{\underline{\beta}}_1, \cdots, \underline{X}_m\tilde{\underline{\beta}}_m) \\ - \quad & \sigma^2 E\left(2\frac{\phi(F)}{F}(mp - p_1) - 4\frac{\phi(F)}{F} - \frac{\phi^2(F)}{F}\right). \end{aligned} \tag{3.23}$$

The corollary follows from (3.23).

Note that under the loss $\sigma^{-2}L(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{a}_1, \cdots, \underline{a}_m)$ $\underline{e}_{LSE}^{*}$ is constant risk minimax estimator of $(\underline{X}\underline{\beta})^*$. Hence, any estimator dominating $\underline{e}_{LSE}^{*}$ is also a minimax estimator of $(\underline{X}\underline{\beta})^*$. Thus, Corollary 3.2.1 provides useful minimax

estimators of $(\underline{X}\underline{\beta})^*$. Also, it shows that $\underline{\varepsilon}^*_{EB}$ dominates $\underline{\varepsilon}^*_{LSE}$, and is a minimax estimator if $\underline{Q}_k = \underline{V}_k^{-1}$, for $\underline{\varepsilon}_{kEB}$ is a special case of $\underline{\varepsilon}_{k\phi}$ with $\phi(F) = mp - p_1 - 2$. It is clear from (3.18) that when $\phi(F) = c$, a positive constant, then $R(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{\varepsilon}_{1\phi}, \cdots, \underline{\varepsilon}_{m\phi})$ is minimized when $c = mp - p_1 - 2$. Thus, for $\underline{Q}_k = \underline{V}_k^{-1}$, $\underline{\varepsilon}^*_{EB}$ is optimal within the class of estimators of the form

$$\underline{X}_{k1}(\hat{\underline{\beta}}_{1.} + (1 - \frac{c}{F})(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})) + (1 - \frac{c}{F})\underline{X}_{k2}\tilde{\underline{\beta}}_{k2},\ k = 1, \cdots, m.$$

It will be shown later that the Bayesian optimality of $\underline{\varepsilon}^*_{EB}$ holds within the same class of estimators irrespective of any p.d. weight matrix $\underline{Q}_k$.

It is not necessarily true that $\underline{\varepsilon}^*_{EB}$ dominate $\underline{\varepsilon}^*_{RLSE} = ((\underline{X}_{11}\hat{\underline{\beta}}_{1.})^T, \cdots, \underline{X}_{m1}\hat{\underline{\beta}}_{1.})^T)^T$. By expressing

$$\begin{aligned} & \underline{X}_{k1}\hat{\underline{\beta}}_{1.} - \underline{X}_k\underline{\beta}_k \\ = \ & \underline{X}_{k1}(\hat{\underline{\beta}}_{1.} - E(\hat{\underline{\beta}}_{1.})) + \underline{X}_{k1}(E(\hat{\underline{\beta}}_{1.}) - E(\hat{\underline{\beta}}_{k1})) \\ + \ & (\underline{X}_{k1}\underline{C}_{k11}^{-1}\underline{C}_{k12} - \underline{X}_{k2})\underline{\beta}_{k2}, \end{aligned}$$

it is easy to check that for estimating $(\underline{X}\underline{\beta})^*$ $\underline{\varepsilon}^*_{RLSE}$ has the risk

$$\begin{aligned} & R(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{X}_{11}\hat{\underline{\beta}}_{1.}, \cdots, \underline{X}_{m1}\hat{\underline{\beta}}_{1.}) \\ = \ & p_1\sigma^2 + \sum_{k=1}^{m} E(\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})^T\underline{C}_{k11}E(\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.}) \\ + \ & \sum_{k=1}^{m} \underline{\beta}_{k2}^T\underline{C}_{k22.1}\underline{\beta}_{k2}. \end{aligned} \tag{3.24}$$

Also, from (3.18) with $\phi(F) = mp - p_1 - 2$, one obtains that

$$\begin{aligned} & R(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{\varepsilon}_{1EB}, \cdots, \underline{\varepsilon}_{mEB}) = mp\sigma^2 \\ - \ & 2\sigma^2(mp - p_1 - 2)^2E(\frac{1}{F}) + (mp - p_1 - 2)^2E(\frac{SSH_1 + SSH_2}{F^2}). \end{aligned} \tag{3.25}$$

For fixed $\underline{\beta}_1, \cdots, \underline{\beta}_m$ and σ^2, SSH_1, SSH_2, and SSE are independently distributed with $(SSH_1 + SSH_2) \sim \sigma^2\,\chi^2_{mp-p_1}(\lambda), \lambda = (1/(2\sigma^2))(\sum_{k=1}^{m} E(\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})^T\,\underline{C}_{k11}\,E(\hat{\underline{\beta}}_{k1} -$

$\hat{\underline{\beta}}_{1.}) + \sum_{k=1}^{m} \underline{\beta}_{k2}^T \underline{C}_{k22.1} \underline{\beta}_{k2})$, and $SSE \sim \sigma^2 \chi^2_{N-mp}$. It follows from (3.25) that

$$\begin{aligned} & R(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{e}_{1EB}, \cdots, \underline{e}_{mEB}) \\ = \; & mp\sigma^2 - \sigma^2(mp - p_1 - 2)^2 (\frac{N - mp}{N - mp + 2}) E(mp - p_1 - 2 + 2K)^{-1}, \quad (3.26) \end{aligned}$$

where $K \sim Poisson(\lambda)$. Thus from (3.24) and (3.26), as $\lambda \to 0$, $R(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{X}_{11}\hat{\underline{\beta}}_{1.}, \cdots, \underline{X}_{m1}\hat{\underline{\beta}}_{1.}) \to p_1\sigma^2$, in which $R(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{e}_{1EB}, \cdots, \underline{e}_{mEB}) \to \sigma^2 \, [\, mp - (mp - p_1 - 2)(N - mp)/(N - mp + 2) \,] > p_1\sigma^2$. On the other hand, as $\lambda \to \infty$, $R(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{X}_{11}\hat{\underline{\beta}}_{1.}, \cdots, \underline{X}_{m1}\hat{\underline{\beta}}_{1.}) \to \infty$, but $R(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{e}_{1EB}, \cdots, \underline{e}_{mEB}) \to mp\sigma^2$. Hence neither $\underline{e}^*_{RLSE}$ nor $\underline{e}^*_{EB}$ dominates each other. The above phenomenon is expected. Small λ indicates that $\underline{\beta}_{k1}$ is close to $\underline{\beta}_{1.}$, and $\underline{\beta}_{k2}$ is close to $\underline{0}$ for all k, in which case clearly $\underline{e}^*_{RLSE}$ is the desired estimator. On the other hand, large λ indicates substantial rejection of $H_0 : \underline{\beta}_{k1} = \underline{\beta}_{1.}$ and $\underline{\beta}_{k2} = \underline{0}$ for all k, in which case the model (3.3) is expected to perform poorly. The EB estimator turns out to be fairly robust. Note that $\underline{e}^*_{RLSE}$ is not a minimax estimator of $(\underline{X}\underline{\beta})^*$. Since $R(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{X}_{11}\hat{\underline{\beta}}_{1.}, \cdots, \underline{X}_{m1}\hat{\underline{\beta}}_{1.}) \to \infty$ as $\lambda \to \infty$, and from the robustness criterion, $\underline{e}^*_{EB}$ is evidently the winner over $\underline{e}^*_{LSE}$ or $\underline{e}^*_{RLSE}$ when $\underline{Q}_k = \underline{V}_k^{-1}$.

However, the estimator is obtained by estimating B, the Bayes shrinking factor, by a random variable which can be bigger than 1 with positive probability. As mentioned earlier, this deficiency is remedied by $\underline{e}^{+*}_{EB}$. For every estimator $\underline{e}^*_\phi$ defined in (3.17). there is a corresponding estimator $\underline{e}^{+*}_\phi$ given by

$$\begin{aligned} \underline{e}^+_{k\phi}(\underline{Y}_k) \; &= \; \underline{X}_{k1}(\hat{\underline{\beta}}_{1.} + (1 - \frac{\phi(F)}{F})^+(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})) \\ &+ \; (1 - \frac{\phi(F)}{F})^+ \underline{X}_{k2}\tilde{\underline{\beta}}_{k2}, \; k = 1, \cdots, m. \end{aligned} \quad (3.27)$$

It can be expected that $\underline{e}^{+*}_\phi$ has smaller risk than $\underline{e}^*_\phi$.

Theorem 3.2.2. Under the loss (3.4) with $\underline{Q}_k = \underline{V}_k^{-1}$,

$$R(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{e}^+_{1\phi}, \cdots, \underline{e}^+_{m\phi}) < R(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{e}_{1\phi}, \cdots, \underline{e}_{m\phi}). \quad (3.28)$$

Proof of Theorem 3.2.2. After some algebra, it can be obtained that

$$
\begin{aligned}
& R(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{\varepsilon}_{1\phi}, \cdots, \underline{\varepsilon}_{m\phi}) - R(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{\varepsilon}^+_{1\phi}, \cdots, \underline{\varepsilon}^+_{m\phi}) \\
= & \sum_{k=1}^{m} E_{\underline{\beta}_1, \cdots, \underline{\beta}_m, \sigma^2}\{[\,(1-\frac{\phi(F)}{F})^2 - ((1-\frac{\phi(F)}{F})^+)^2\,] \\
& (\underline{X}_{k1}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.}) + \underline{X}_{k2}\tilde{\underline{\beta}}_{k2})^T \underline{V}_k^{-1} (\underline{X}_{k1}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.}) + \underline{X}_{k2}\tilde{\underline{\beta}}_{k2}) \\
+ & 2[\,(1-\frac{\phi(F)}{F}) - (1-\frac{\phi(F)}{F})^+\,] \\
& (\underline{X}_{k1}\hat{\underline{\beta}}_{1.} - \underline{X}_k\underline{\beta}_k)^T \underline{V}_k^{-1} (\underline{X}_{k1}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.}) + \underline{X}_{k2}\tilde{\underline{\beta}}_{k2})\} \\
\geq & \sum_{k=1}^{m} -2E_{\underline{\beta}_1, \cdots, \underline{\beta}_m, \sigma^2}\{(1-\frac{\phi(F)}{F}) I_{[\frac{\phi(F)}{F} > 1]} \\
& (\underline{X}_{k1}\hat{\underline{\beta}}_{1.} - \underline{X}_k\underline{\beta}_k)^T \underline{V}_k^{-1} (\underline{X}_{k1}\hat{\underline{\beta}}_{1.} - \underline{X}_k\tilde{\underline{\beta}}_k)\}. \qquad (3.29)
\end{aligned}
$$

From (3.29), it is suffices to show that

$$
\begin{aligned}
& \sum_{k=1}^{m} E_{\underline{\beta}_1, \cdots, \underline{\beta}_m, \sigma^2}\{(\underline{X}_{k1}\hat{\underline{\beta}}_{1.} - \underline{X}_k\underline{\beta}_k)^T \underline{V}_k^{-1} (\underline{X}_{k1}\hat{\underline{\beta}}_{1.} - \underline{X}_k\tilde{\underline{\beta}}_k) \\
& |(\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})^T \underline{C}_{k11} (\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.}) = h_{1k},\ \tilde{\underline{\beta}}^T_{k2} \underline{C}_{k22.1} \tilde{\underline{\beta}}_{k2} = h_{2k},\ \forall k, \\
& SSE = e,\ \exists \frac{\phi(F)}{F} > 1\} > 0, \qquad (3.30)
\end{aligned}
$$

for all $h_{1k} > 0$, $h_{2k} > 0$ and $e > 0$. Next, using

$$
\begin{aligned}
& \underline{X}_{k1}\hat{\underline{\beta}}_{1.} - \underline{X}_k\underline{\beta}_k = \underline{X}_{k1}(\hat{\underline{\beta}}_{1.} - E(\hat{\underline{\beta}}_{1.})) \\
+ & \underline{X}_{k1}(E(\hat{\underline{\beta}}_{1.}) - E(\hat{\underline{\beta}}_{k1})) + (\underline{X}_{k1}\underline{C}^{-1}_{k11}\underline{C}_{k12} - \underline{X}_{k2})\underline{\beta}_{k2}, \qquad (3.31)
\end{aligned}
$$

and

$$
\begin{aligned}
& \underline{X}_{k1}\hat{\underline{\beta}}_{1.} - \underline{X}_k\tilde{\underline{\beta}}_k \\
= & \underline{X}_{k1}(\hat{\underline{\beta}}_{1.} - \hat{\underline{\beta}}_{k1}) + (\underline{X}_{k1}\underline{C}^{-1}_{k11}\underline{C}_{k12} - \underline{X}_{k2})\tilde{\underline{\beta}}_{k2}, \qquad (3.32)
\end{aligned}
$$

Eq.(3.30) can be expressed as

$$
\begin{aligned}
& \sum_{k=1}^{m} E_{\underline{\beta}_1, \cdots, \underline{\beta}_m, \sigma^2}\{(E(\hat{\underline{\beta}}_{1.} - \hat{\underline{\beta}}_{k1}))^T \underline{C}_{k11} (\hat{\underline{\beta}}_{1.} - \hat{\underline{\beta}}_{k1}) \\
+ & \underline{\beta}^T_{k2}\underline{C}_{k22.1}\tilde{\underline{\beta}}_{k2} | (\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})^T \underline{C}_{k11} (\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.}) = h_{1k}, \\
& \tilde{\underline{\beta}}^T_{k2}\underline{C}_{k22.1}\tilde{\underline{\beta}}_{k2} = h_{2k},\ \forall k,\ SSE = e,\ \exists \frac{\phi(F)}{F} > 1\}. \qquad (3.33)
\end{aligned}
$$

Since $(\hat{\underline{\beta}}_{11}^T \cdots \hat{\underline{\beta}}_{m1}^T)$ $(\tilde{\underline{\beta}}_{12}^T \cdots \tilde{\underline{\beta}}_{m2}^T)$ and SSE are mutually independent, (3.33) can be written as

$$\begin{aligned}
&\sum_{k=1}^{m} E_{\underline{\beta}_{\cdot 1},\cdots,\underline{\beta}_{\cdot m},\sigma^2}\{(E(\hat{\beta}_{1\cdot} - \hat{\underline{\beta}}_{k1})^T \underline{C}_{k11}(\hat{\beta}_{1\cdot} - \hat{\underline{\beta}}_{k1}) \\
&|(\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{\cdot 1})^T \underline{C}_{k11}(\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{\cdot 1}) = h_{1k}, \forall k) \\
+ \quad &(\underline{\beta}_{k2}^T \underline{C}_{k22.1} \tilde{\underline{\beta}}_{k2} | \tilde{\underline{\beta}}_{k2}^T \underline{C}_{k22.1} \tilde{\underline{\beta}}_{k2} = h_{2k})\}. \qquad (3.34)
\end{aligned}$$

From the proof of Theorem 2.2.3, it can be shown that

$$\begin{aligned}
&\sum_{k=1}^{m} E_{\underline{\beta}_k,\sigma^2}\{(\underline{\beta}_{k2}^T \underline{C}_{k22.1} \tilde{\underline{\beta}}_{k2} | \tilde{\underline{\beta}}_{k2}^T \underline{C}_{k22.1} \tilde{\underline{\beta}}_{k2} = h_{2k}\} > 0, \qquad (3.35) \\
&\forall h_{2k} > 0.
\end{aligned}$$

Recall $\underline{C} = \sum_{k=1}^{m} \underline{C}_{k11}$,

$$\underline{A} = \begin{bmatrix} \underline{C}_{111} & \cdots & \underline{0} \\ \vdots & \ddots & \vdots \\ \underline{0} & \cdots & \underline{C}_{m11} \end{bmatrix} - \begin{bmatrix} \underline{C}_{111}\underline{C}^{-1}\underline{C}_{111} & \cdots & \underline{C}_{111}\underline{C}^{-1}\underline{C}_{m11} \\ \vdots & \ddots & \vdots \\ \underline{C}_{m11}\underline{C}^{-1}\underline{C}_{111} & \cdots & \underline{C}_{m11}\underline{C}^{-1}\underline{C}_{m11} \end{bmatrix},$$

and $SSH_1 = \sum_{k=1}^{m} (\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{\cdot 1})^T \underline{C}_{k11}(\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{\cdot 1})$

$$= [\hat{\underline{\beta}}_{11}^T \cdots \hat{\underline{\beta}}_{m1}^T] \underline{A} \begin{bmatrix} \hat{\underline{\beta}}_{11} \\ \vdots \\ \hat{\underline{\beta}}_{m1} \end{bmatrix}.$$

Since

$$V\left(\begin{bmatrix} \hat{\underline{\beta}}_{11} \\ \vdots \\ \hat{\underline{\beta}}_{m1} \end{bmatrix}\right) = \begin{bmatrix} \underline{C}_{111}^{-1} & \cdots & \underline{0} \\ \vdots & \ddots & \vdots \\ \underline{0} & \cdots & \underline{C}_{m11}^{-1} \end{bmatrix}$$

is p.d., there exists a nonsingular $\underline{Z}$ such that $\underline{V} = \underline{Z}\,\underline{Z}^T$. Now,

$$SSH_1 = \underline{\phi}^T \underline{Z}^T \underline{A} \underline{Z} \underline{\phi} \qquad (3.36)$$

with $\underline{Z}^T \underline{A} \underline{Z}$ being idempotent and $rank(\underline{Z}^T \underline{A} \underline{Z}) = (m-1)p_1$, and

$$\underline{\phi} = \underline{Z}^{-1} \begin{bmatrix} \hat{\underline{\beta}}_{11} \\ \vdots \\ \hat{\underline{\beta}}_{m1} \end{bmatrix} \sim N_{mp_1}(\underline{Z}^{-1} E \begin{bmatrix} \hat{\underline{\beta}}_{11} \\ \vdots \\ \hat{\underline{\beta}}_{m1} \end{bmatrix}, \underline{I}_{mp_1}).$$

There exists an orthogonal matrix $\underline{P}$ such that

$$\underline{P}^T(\underline{Z}^T\underline{A}\,\underline{Z})\underline{P} = \begin{bmatrix} \underline{I}_{(m-1)p_1} & \underline{0}_{(m-1)p_1\times p_1} \\ \underline{0}_{p_1\times(m-1)p_1} & \underline{0}_{p_1\times p_1} \end{bmatrix}. \tag{3.37}$$

From (3.36),

$$SSH_1 = (\underline{P}\underline{\phi})^T\underline{P}(\underline{Z}^T\underline{A}\,\underline{Z})\underline{P}^T(\underline{P}\underline{\phi}), \tag{3.38}$$

note that

$$\underline{\alpha} = \underline{P}\,\underline{\phi} \sim N_{mp_1}(\underline{P}\underline{Z}^{-1}E\begin{bmatrix} \hat{\underline{\beta}}_{11} \\ \vdots \\ \hat{\underline{\beta}}_{m1} \end{bmatrix}, \underline{I}_{mp_1}), \tag{3.39}$$

combining (3.37) (3.38) and (3.39), one gets

$$\begin{aligned} SSH_1 &= \Sigma_{i=1}^{(m-1)p_1}\alpha_i^2 \\ &= \underline{\alpha}_{(1)}^T\underline{\alpha}_{(1)}, \end{aligned} \tag{3.40}$$

where α_i is the ith element of $\underline{\alpha}$, and $\underline{\alpha}_{(1)}^T = (\alpha_1 \cdots \alpha_{(m-1)p_1})$. Using (3.39) and (3.40), and applying the argument in Theorem 6.2 of Lehmann (1983), yields

$$\begin{aligned} &\sum_{k=1}^{m} E_{\underline{\beta}_{1\cdot},\cdots,\underline{\beta}_{m\cdot},\sigma^2}\{(E(\hat{\underline{\beta}}_{1\cdot} - \underline{\beta}_{k1}))^T\underline{C}_{k11}(\hat{\underline{\beta}}_{1\cdot} - \underline{\beta}_{k1})|SSH_1 = h_1\} \\ = \; & E_{\underline{\beta}_{1\cdot},\cdots,\underline{\beta}_{m\cdot},\sigma^2}\{(E(\underline{\alpha}_{(1)}))^T\underline{\alpha}_{(1)}|\underline{\alpha}_{(1)}^T\underline{\alpha}_{(1)}\} \geq 0. \end{aligned} \tag{3.41}$$

Theorem 3.2.2 is hence proved as a consequence from (3.35) and (3.41).

Finally, based on the hypothesis $H_0 : \underline{\beta}_{k1} = \underline{\beta}_{1\cdot}$ and $\underline{\beta}_{k2} = \underline{0}$ for all k, a general PTE is given by

$$\underline{\delta}_k^{PTE} = \underline{X}_{k1}(\hat{\underline{\beta}}_{1\cdot} + g(F)(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1\cdot})) + \underline{X}_{k2}g(F)\tilde{\underline{\beta}}_{k2}, \forall\, k, \tag{3.42}$$

where $g(F) = I_{[F>d]}$ with d a postive constant depending on the chosen level of significance, and I the usual indicator function. The corresponding modified EB estimator,

$$\begin{aligned} & \underline{\varepsilon}_{kMEB}^{\phi_0}(\underline{Y}_k) \\ = \; & \underline{X}_{k1}(\hat{\underline{\beta}}_{1\cdot} + (1 - \frac{\phi_0(F)}{F})g(F)(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1\cdot})) \\ + \; & \underline{X}_{k2}(1 - \frac{\phi_0(F)}{F})g(F)\tilde{\underline{\beta}}_{k2}, \; k = 1, \cdots, m, \end{aligned} \tag{3.43}$$

dominates $\underline{\delta}^*_{PTE} = ((\underline{\varepsilon}_1^{PTE})^T, \cdots, (\underline{\varepsilon}_m^{PTE})^T)^T$ under certain conditions. This is shown in the following theorem.

Theorem 3.2.3. Consider the loss (3.4) with $\underline{Q}_k = \underline{V}_k^{-1}$. If $0 < \phi_0(F) < 2(mp - p_1 - 2)$, then

$$R(\underline{X}_1\underline{\beta}_{.1}, \cdots, \underline{X}_m\underline{\beta}_{.m}; \underline{\delta}^{\phi_0}_{1MEB}, \cdots, \underline{\delta}^{\phi_0}_{mMEB}) < R(\underline{X}_1\underline{\beta}_{.1}, \cdots, \underline{X}_m\underline{\beta}_{.m}; \underline{\delta}_1^{PTE}, \cdots, \underline{\delta}_m^{PTE}). \tag{3.44}$$

Proof of Theorem 3.2.3. Using (3.18), and following the procedure in Proof of Theorem 2.2.4, (3.44) can be obtained immediately.

Note that $\underline{\delta}^{\phi_0}_{kMEB}$ estimates $\underline{X}_k\underline{\beta}_{.k}$ by $\underline{X}_{k1}\underline{\hat{\beta}}_{.1}$ if $g(F) = 0$, i.e., the null hypothesis is accepted at a desired level of significance, and it estimates $\underline{X}_k\underline{\beta}_{.k}$ by an EB type estimator as described earlier in this section when the other is true. Also, the modified EB estimator $\underline{\delta}^{\phi_0}_{MEB} = ((\underline{\delta}^{\phi_0}_{1MEB})^T, \cdots, (\underline{\delta}^{\phi_0}_{mMEB})^T)^T$ can be further improved by a positive part modified estimator $(\underline{\delta}^{\phi_0}_{MEB})^{+*} = ((\underline{\delta}^{\phi_0^+}_{1MEB})^T, \cdots, (\underline{\delta}^{\phi_0^+}_{mMEB})^T)^T$ as

$$\begin{aligned} & \underline{\delta}^{\phi_0^+}_{kMEB}(\underline{Y}_k) \\ = \; & \underline{X}_{k1}(\underline{\hat{\beta}}_{.1} + (1 - \frac{\phi_0(F)}{F})^+ g(F)(\underline{\tilde{\beta}}_{k1} - \underline{\hat{\beta}}_{.1})) \\ + \; & \underline{X}_{k2}(1 - \frac{\phi_0(F)}{F})^+ g(F)\underline{\tilde{\beta}}_{k2}, k = 1, \cdots, m, \end{aligned} \tag{3.45}$$

Now, the $\underline{\varepsilon}^*_{EB}$ and other associated estimators of $(\underline{X}\underline{\beta})^*$ are evaluated by the Bayes risk. In view of Theorem 3.2.1, the Bayes risk of $\underline{\varepsilon}^*_{EB}$ is smaller than that of $\underline{\varepsilon}^*_{LSE}$ when $\underline{Q}_k = \underline{V}_k^{-1}$ without regarding of any prior, while $\underline{\varepsilon}^{+*}_{EB}$ always dominates $\underline{\varepsilon}^*_{EB}$ irrespective of any prior. However, the Bayes risk optimatity of $\underline{\varepsilon}^*_{EB}$ over $\underline{\varepsilon}^*_{LSE}$ is not limited to $\underline{Q}_k = \underline{V}_k^{-1}$. The following theorem will show this superiority of Bayes risk of $\underline{\varepsilon}^*_{EB}$ over $\underline{\varepsilon}^*_{LSE}$ under the prior presented in the beginning of this section.

Theorem 3.2.4. Consider the model $\underline{Y}_k|\underline{\beta}_{.k} \sim N(\underline{X}_k\underline{\beta}_{.k}, \sigma^2\underline{V}_k)$, for $k = 1, \cdots, m$, where $\underline{V}_k$'s are p.d. and $\underline{\beta}_{.k} \sim N(\underline{\nu}, \tau^2(\underline{X}_k^T\underline{V}_k^{-1}\underline{X}_k)^{-1})$ for all k, and $\underline{\nu}^T = (\underline{\nu}_1^T \; \underline{0}^T)$.

Let ξ denote this given prior and let $r(\xi;\underline{e}_1,\cdots,\underline{e}_m)$ denote the Bayes risk of an estimator $(\underline{e}_1,\cdots,\underline{e}_m)$ of $(\underline{X}_1\underline{\beta}_1,\cdots,\underline{X}_m\underline{\beta}_m)$ under the prior ξ when the loss is given in (3.4). Then,

$$r(\xi;\underline{e}_{1EB},\cdots,\underline{e}_{mEB}) < r(\xi;\underline{X}_1\tilde{\underline{\beta}}_1,\cdots,\underline{X}_m\tilde{\underline{\beta}}_m). \tag{3.46}$$

Proof of Theorem 3.2.4. Let

$$\underline{e}_{kc} = \underline{X}_{k1}(\tilde{\underline{\beta}}_{k1} - \frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})) + (1 - \frac{c}{F})\underline{X}_{k2}\tilde{\underline{\beta}}_{k2},\ k = 1,\cdots,m. \tag{3.47}$$

According to the loss (3.4),

$$\begin{aligned}
& r(\xi;\underline{e}_{1c},\cdots,\underline{e}_{mc}) \\
= & \sum_{k=1}^{m} tr[\,\{E(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1} - \frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})) \\
& (\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1} - \frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.}))^T\}\underline{X}_{k1}^T\underline{Q}_k\underline{X}_{k1}\,] \\
+ & \sum_{k=1}^{m} tr[\,\{E(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1} - \frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})) \\
& (\tilde{\underline{\beta}}_{k2} - \underline{\beta}_{k2} - \frac{c}{F}\tilde{\underline{\beta}}_{k2})^T\}\underline{X}_{k2}^T\underline{Q}_k\underline{X}_{k1}\,] \\
+ & \sum_{k=1}^{m} tr[\,\{E(\tilde{\underline{\beta}}_{k2} - \underline{\beta}_{k2} - \frac{c}{F}\tilde{\underline{\beta}}_{k2}) \\
& (\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1} - \frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.}))^T\}\underline{X}_{k1}^T\underline{Q}_k\underline{X}_{k2}\,] \\
+ & \sum_{k=1}^{m} tr[\,E(\tilde{\underline{\beta}}_{k2} - \underline{\beta}_{k2} - \frac{c}{F}\tilde{\underline{\beta}}_{k2})(\tilde{\underline{\beta}}_{k2} - \underline{\beta}_{k2} - \frac{c}{F}\tilde{\underline{\beta}}_{k2})^T\}\underline{X}_{k2}^T\underline{Q}_k\underline{X}_{k2}\,].
\end{aligned} \tag{3.48}$$

Since

$$\begin{aligned}
& E(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1} - \frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.}))(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1} - \frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.}))^T \\
= & \ \sigma^2\underline{Q}_{k11.2}^{-1} + E(\frac{c^2}{F^2}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})^T) \\
- & \ E(\frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1})(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})^T) - E(\frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1})^T),
\end{aligned} \tag{3.49}$$

and

$$\begin{aligned}
& E(\frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1})(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})^T) \\
= & \ \sigma^2\frac{c(N-mp)}{N-mp+2}E(\frac{1}{SSH_1 + SSH_2}(\tilde{\underline{\beta}}_{k1} - E(\underline{\beta}_{k1}|\underline{Y}_k))(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})^T).
\end{aligned} \tag{3.50}$$

By using (3.6), (3.50) can be expressed as

$$\sigma^2 B\frac{c(N-mp)}{N-mp+2}E(\frac{1}{SSH_1+SSH_2}\{(\tilde{\underline{\beta}}_{k1}-\hat{\underline{\beta}}_{.1})(\tilde{\underline{\beta}}_{k1}-\hat{\underline{\beta}}_{.1})^T + (\hat{\underline{\beta}}_{.1}-\underline{\nu}_1)(\tilde{\underline{\beta}}_{k1}-\hat{\underline{\beta}}_{.1})^T\}). \tag{3.51}$$

Using (3.11), independence between $\hat{\underline{\beta}}_{k1}$'s and $\tilde{\underline{\beta}}_{k2}$'s, independence between $\hat{\underline{\beta}}_{.1}$ and SSH_1, independence between $\hat{\underline{\beta}}_{.1}$ and $\hat{\underline{\beta}}_{k1}-\hat{\underline{\beta}}_{.1}$, and $E\hat{\underline{\beta}}_{.1} = E\hat{\underline{\beta}}_{k1} = \underline{\nu}_1$ for all k, applying Basu's theorem, and Lemma1 in Ghosh, Saleh, and Sen (1987), Equation (3.51) equals to

$$\begin{aligned}&\sigma^2 B\frac{c(N-mp)}{N-mp+2}\{\frac{V(\hat{\underline{\beta}}_{k1}-\hat{\underline{\beta}}_{.1})}{E(SSH_1+SSH_2)} \\ &+ \frac{\underline{C}_{k11}^{-1}\underline{C}_{k12}V(\tilde{\underline{\beta}}_{k2})\underline{C}_{k21}\underline{C}_{k11}^{-1}}{E(SSH_1+SSH_2)}\} \\ &= \sigma^2 B\frac{c(N-mp)}{(N-mp+2)}\frac{1}{(mp-p_1)}(\underline{C}_{k11.2}^{-1}-\underline{C}^{-1}),\end{aligned} \tag{3.52}$$

where, $\underline{C}=\sum_{k=1}^{m}\underline{C}_{k11}$. Again, according to the same argument in (3.52),

$$\begin{aligned}&E(\frac{c^2}{F^2}(\tilde{\underline{\beta}}_{k1}-\hat{\underline{\beta}}_{.1})(\tilde{\underline{\beta}}_{k1}-\hat{\underline{\beta}}_{.1})^T) \\ &= \sigma^2 B\frac{c^2(N-mp)}{(N-mp+2)}\frac{(\underline{C}_{k11.2}^{-1}-\underline{C}^{-1})}{(mp-p_1)(mp-p_1-2)}.\end{aligned} \tag{3.53}$$

And combining (3.49), (3.52) and (3.53), the first term of (3.48) can be expressed by

$$\begin{aligned}&\sum_{k=1}^{m}\{\sigma^2 tr[\,\underline{C}_{k11.2}^{-1}(\underline{X}_{k1}^T\underline{Q}_k\underline{X}_{k1})+(\frac{c^2}{mp-p_1-2}-2c)\sigma^2 B\frac{N-mp}{(N-mp+2)(mp-p_1)} \\ &tr[\,(\underline{C}_{k11.2}^{-1}-\underline{C}^{-1})\underline{X}_{k1}^T\underline{Q}_k\underline{X}_{k1}\,]\}.\end{aligned} \tag{3.54}$$

Now, applying Basu's theorem and Lemma 1 in Ghosh, Saleh, and Sen (1987),

$$\begin{aligned}&E(\tilde{\underline{\beta}}_{k2}-\underline{\beta}_{k2}-\frac{c}{F}\tilde{\underline{\beta}}_{k2})(\tilde{\underline{\beta}}_{k2}-\underline{\beta}_{k2}-\frac{c}{F}\tilde{\underline{\beta}}_{k2})^T \\ &= \sigma^2\underline{C}_{k22.1}^{-1}-E(\frac{c}{F}\tilde{\underline{\beta}}_{k2}(\tilde{\underline{\beta}}_{k2}-E(\underline{\beta}_{k2}|\underline{Y}_k))^T)+E(\frac{c^2}{F^2}\tilde{\underline{\beta}}_{k2}\tilde{\underline{\beta}}_{k2}^T) \\ &= \sigma^2\underline{C}_{k22.1}^{-1}+(\frac{c^2}{mp-p_1-2}-2c)\sigma^2 B\frac{N-mp}{(N-mp+2)(mp-p_1)}\underline{C}_{k22.1}^{-1}.\end{aligned} \tag{3.55}$$

From (3.55), the last term of (3.48) becomes

$$\sum_{k=1}^{m}\{\sigma^2 tr\,[\,\underline{C}_{k22.1}^{-1}(\underline{X}_{k2}^T\underline{Q}_k\underline{X}_{k2})\,] + (\frac{c^2}{mp-p_1-2} - 2c)\sigma^2 B\frac{N-mp}{(N-mp+2)(mp-p_1)}$$
$$tr\,[\,\underline{C}_{k22.1}^{-1}\underline{X}_{k2}^T\underline{Q}_k\underline{X}_{k2}\,]\}. \tag{3.56}$$

Finally, by using the same argument, the cross-product term can be expressed as

$$\begin{aligned}
& E(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1} - \frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{.1}))(\tilde{\underline{\beta}}_{k2} - \underline{\beta}_{k2} - \frac{c}{F}\tilde{\underline{\beta}}_{k2})^T \\
= \;& -\sigma^2(\underline{C}_{k11.2}^{-1}\underline{C}_{k12}\underline{C}_{k22}^{-1}) - E(\frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{.1})(\tilde{\underline{\beta}}_{k2} - E(\underline{\beta}_{k2}|\underline{Y}_k))^T) \\
- \;& E(\frac{c}{F}(\tilde{\underline{\beta}}_{k1} - E(\underline{\beta}_{k1}|\underline{Y}_k))\tilde{\underline{\beta}}_{k2}^T) + E(\frac{c^2}{F^2}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{.1})\tilde{\underline{\beta}}_{k2}^T) \\
= \;& -\sigma^2(\underline{C}_{k11.2}^{-1}\underline{C}_{k12}\underline{C}_{k22}^{-1}) - 2\sigma^2 B\frac{c(N-mp)}{N-mp-2}E(\frac{-\underline{C}_{k11}^{-1}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2}\tilde{\underline{\beta}}_{k2}^T}{SSH_1 + SSH_2}) \\
+ \;& \sigma^4\frac{c^2(N-mp)}{N-mp+2}E(\frac{-\underline{C}_{k11}^{-1}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2}\tilde{\underline{\beta}}_{k2}^T}{SSH_1 + SSH_2})^2 \\
= \;& -\sigma^2(\underline{C}_{k11.2}^{-1}\underline{C}_{k12}\underline{C}_{k22}^{-1}) \\
- \;& (\frac{c^2}{mp-p_1-2} - 2c)\sigma^2 B\frac{N-mp}{(N-mp+2)(mp-p_1)}\underline{C}_{k11}^{-1}\underline{C}_{k12}\underline{C}_{k22.1}^{-1}. \qquad (3.57)
\end{aligned}$$

From (3.57), the cross-product term in (3.48) can be expressed as

$$\begin{aligned}
& \sum_{k=1}^{m}\{-\sigma^2 tr\,[\,\underline{C}_{k11.2}^{-1}\underline{C}_{k12}\underline{C}_{k22}^{-1}\underline{X}_{k2}^T\underline{Q}_k\underline{X}_{k1}\,] \\
- \;& (\frac{c^2}{mp-p_1-2} - 2c)\sigma^2 B\frac{N-mp}{(N-mp+2)(mp-p_1)}tr\,[\,\underline{C}_{k11}^{-1}\underline{C}_{k12}\underline{C}_{k22.1}^{-1}\underline{X}_{k2}^T\underline{Q}_k\underline{X}_{k1}\,] \\
- \;& \sigma^2 tr\,[\,\underline{C}_{k22.1}^{-1}\underline{C}_{k21}\underline{C}_{k11}^{-1}\underline{X}_{k1}^T\underline{Q}_k\underline{X}_{k2}\,] - (\frac{c^2}{mp-p_1-2} - 2c)\sigma^2 B\frac{N-mp}{(N-mp+2)(mp-p_1)} \\
& tr\,[\,\underline{C}_{k22}^{-1}\underline{C}_{k21}\underline{C}_{k11.2}^{-1}\underline{X}_{k1}^T\underline{Q}_k\underline{X}_{k2}\,]\}. \qquad (3.58)
\end{aligned}$$

By combining (3.54), (3.55) and (3.58), it can be obtained that

$$\begin{aligned}
& r(\xi; \underline{e}_{1c}, \cdots, \underline{e}_{mc}) \\
= \;& \sum_{k=1}^{m}\{\sigma^2 tr\,[\,(\underline{X}_k^T\underline{V}_k^{-1}\underline{X}_k)^{-1}\underline{X}_k^T\underline{Q}_k\underline{X}_k\,] \\
+ \;& (\frac{c^2}{mp-p_1-2} - 2c)\sigma^2 B\frac{N-mp}{(N-mp+2)(mp-p_1)}(tr\,[\,\underline{G}_k\underline{C}_{k22.1}^{-1}\,] \\
+ \;& tr\,[\,(\underline{C}_{k11}^{-1} - \underline{C}^{-1})\underline{X}_{k1}^T\underline{Q}_k\underline{X}_{k1}\,]\}, \qquad (3.59)
\end{aligned}$$

where $\underline{G}_k = (\underline{X}_{k1}\underline{C}_{11}^{-1}\underline{C}_{k12} - \underline{X}_{k2})^T\underline{Q}_k\ (\underline{X}_{k1}\underline{C}_{11}^{-1}\underline{C}_{k12} - \underline{X}_{k2})$. From (3.59), it can be concluded that $r(\xi; \underline{e}_{1c}, \cdots, \underline{e}_{mc}) < r(\xi; \underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m)$, when $0 < c < 2(mp - p_1 - 2)$. Since $\underline{e}_{kc} = \underline{e}_{kEB}$ if $c = mp - p_1 - 2$, Theorem 3.2.4 is proved.

The next theorem gives necessary and sufficient conditions under which the Bayes risk of $\underline{e}_{EB}^*$ is smaller than that of $\underline{e}_{RLSE}^*$.

Theorem 3.2.5. Consider the set of Theorem 3.2.4. Then,

$$r(\xi, \underline{e}_{1EB}, \cdots, \underline{e}_{mEB}) < r(\xi, \underline{X}_{11}\hat{\underline{\beta}}_{1.}, \cdots, \underline{X}_{m1}\hat{\underline{\beta}}_{1.}), \tag{3.60}$$

if and only if $(1-B)^2/B^2 > (1-(mp-p_1-2)(N-mp)/((mp-p_1)(N-mp+2)))$ with $B = \sigma^2/(\sigma^2+\tau^2)$.

Proof of Theorem 3.2.5. By expressing

$$\begin{aligned}
& \underline{X}_{k1}\hat{\underline{\beta}}_{1.} - \underline{X}_k\underline{\beta}_k \\
= \ & \underline{X}_{k1}(\hat{\underline{\beta}}_{1.} - E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}(\hat{\underline{\beta}}_{1.})) \\
+ \ & \underline{X}_{k1}(E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}(\hat{\underline{\beta}}_{1.}) - E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}(\hat{\underline{\beta}}_{k1})) \\
+ \ & (\underline{X}_{k1}\underline{C}_{11}^{-1}\underline{C}_{k12} - \underline{X}_{k2})\underline{\beta}_{k2},
\end{aligned}$$

it can be checked that

$$\begin{aligned}
& R(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{X}_{11}\hat{\underline{\beta}}_{1.}, \cdots, \underline{X}_{m1}\hat{\underline{\beta}}_{1.}) \\
= \ & \sum_{k=1}^{m}\{tr\,[\,V(\hat{\underline{\beta}}_{1.})(\underline{X}_{k1}^T\underline{Q}_k\underline{X}_{k1})\,] \\
+ \ & tr\,[\,E(\hat{\underline{\beta}}_{1.} - \hat{\underline{\beta}}_{k1})E(\hat{\underline{\beta}}_{1.} - \hat{\underline{\beta}}_{k1})^T(\underline{X}_{k1}^T\underline{Q}_k\underline{X}_{k1})\,] \\
+ \ & 2tr\,[\,\underline{\beta}_{k2}E(\hat{\underline{\beta}}_{1.} - \hat{\underline{\beta}}_{k1})^T(\underline{X}_{k1}^T\underline{Q}_k(\underline{X}_{k1}\underline{C}_{11}^{-1}\underline{C}_{k12} - \underline{X}_{k2}))\,]\,] \\
+ \ & tr\,[\underline{\beta}_{k2}\underline{\beta}_{k2}^T\underline{G}_k\,]\},
\end{aligned} \tag{3.61}$$

where $\underline{G}_k = (\underline{X}_{k1}\underline{C}_{11}^{-1}\underline{C}_{k12} - \underline{X}_{k2})^T\underline{Q}_k\ (\underline{X}_{k1}\underline{C}_{11}^{-1}\underline{C}_{k12} - \underline{X}_{k2})$. Since

$$\underline{\beta}_k \sim^{ind} N_p\Big(\begin{pmatrix}\underline{\nu}_1\\ \underline{0}\end{pmatrix}, \tau^2(\underline{X}_k^T\underline{V}_k^{-1}\underline{X}_k)^{-1}\Big), \tag{3.62}$$

$$
\begin{aligned}
& E_{\underline{\nu}_1,\tau^2}(\underline{\beta}_{k2}\underline{E}_{\underline{\beta}_1,\cdots,\underline{\beta}_m}(\hat{\underline{\beta}}_{1.} - \hat{\underline{\beta}}_{k1})^T) \\
= \; & Cov(\underline{\beta}_{k2}, (\underline{C}^{-1}\underline{C}_{k11} - I)(\underline{\beta}_{k1} + \underline{C}_{k11}^{-1}\underline{C}_{k12}\underline{\beta}_{k2})) \\
= \; & \underline{0}. \qquad (3.63)
\end{aligned}
$$

By using (3.61),(3.62) and (3.63), one gets

$$
\begin{aligned}
& r(\xi, \underline{X}_{11}\hat{\underline{\beta}}_{1.}, \cdots, \underline{X}_{m1}\hat{\underline{\beta}}_{1.}) \\
= \; & \sigma^2 \sum_{k=1}^{m} \{tr\,[\,\underline{C}^{-1}(\underline{X}_{k1}^T\underline{Q}_k\underline{X}_{k1}\,] \\
+ \; & tr\,[\,V_{\underline{\nu}_1,\sigma^2}(\underline{E}_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}(\hat{\underline{\beta}}_{1.} - \hat{\underline{\beta}}_{k1}))(\underline{X}_{k1}^T\underline{Q}_k\underline{X}_{k1})\,] \\
+ \; & \tau^2 tr\,[\underline{G}_k\underline{C}_{k22.1}^{-1}]\}. \qquad (3.64)
\end{aligned}
$$

Now, $V_{\underline{\nu}_1,\sigma^2}(\underline{E}_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\,(\hat{\underline{\beta}}_{1.} - \hat{\underline{\beta}}_{k1})) = \tau^2(\underline{C}_{k11}^{-1} - \underline{C}^{-1})$. By using $tr\,[\,(\underline{X}_k^T\underline{V}_k^{-1}\underline{X}_k)^{-1}\,(\underline{X}_k^T\underline{Q}_k\underline{X}_k)\,] = tr\,[\,\underline{C}_{k11}^{-1}\,(\underline{X}_{k1}^T\underline{Q}_k\underline{X}_{k1})\,] + tr\,[\underline{G}_k\underline{C}_{k22.1}^{-1}]$, (3.59) with $c = mp - p_1 - 2$ and (3.64),

$$
\begin{aligned}
& r(\xi, \underline{X}_{11}\hat{\underline{\beta}}_{1.}, \cdots, \underline{X}_{m1}\hat{\underline{\beta}}_{1.}) - r(\xi, \underline{e}_{1EB}, \cdots, \underline{e}_{mEB}) \\
= \; & (\tau^2 - \sigma^2 + \sigma^2 B\frac{(N-mp)}{(N-mp+2)}\frac{(mp-p_1-2)}{(mp-p_1)})(\sum_{k=1}^{m}\{tr\,[\,\underline{C}_{k22.1}^{-1}\underline{G}_k\,] \\
+ \; & tr\,[\,(\underline{C}_{k11}^{-1} - \underline{C}^{-1})\underline{X}_{k1}^T\underline{Q}_k\underline{X}_{k1}\,]\}). \qquad (3.65)
\end{aligned}
$$

From (3.65), the conclusion of the theorem follows.

The consequence of the above theorem is clear. If $\sigma^2 < \tau^2$, then $B < 1/2$ in which case $(1-B)^2/B^2 > 1$. Thus, $\underline{e}_{EB}^*$ always dominates $\underline{e}_{RLSE}^*$ in its Bayes risk if the sampling variability is smaller than the prior variability. However, if τ^2 is much smaller than σ^2, the model (3.3) seems to be appropriate, and clearly $\underline{e}_{RLSE}^*$ becomes a winner.

A general class of minimax estimates of $(\underline{X\beta})^*$, which was described earlier in this section, can be expressed as

$$
\begin{aligned}
& \underline{e}_{kc}(\underline{Y}_k) \\
= \; & \underline{X}_{k1}(\hat{\underline{\beta}}_{1.} + (1 - \frac{c}{F})(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})) + \underline{X}_{k2}(1 - \frac{c}{F})\tilde{\underline{\beta}}_{k2},\; k = 1, \cdots, m, \quad (3.66)
\end{aligned}
$$

where c is a proper constant. The following theorem shows that $\underline{e}^*_{EB}$ is the optimal estimate of $(\underline{X\beta})^*$ within the class of all estimators of the form $\underline{e}^*_c = (\underline{e}^T_{1c}, \cdots, \underline{e}^T_{mc})^T$.

Theorem 3.2.6. Consider the same set up as of Theorem 3.2.4. Then,

$$\begin{aligned} & r(\xi, \underline{e}_{1c}, \cdots, \underline{e}_{mc}) - r(\xi, \underline{e}_{1B}, \cdots, \underline{e}_{mB}) \\ = \; & \sigma^2 B \sum_{k=1}^{m} tr\,[(\underline{X}_k^T \underline{V}_k^{-1} \underline{X}_k)^{-1} \underline{X}_k^T \underline{Q}_k \underline{X}_k] \\ + \; & (\frac{c^2}{mp - p_1 - 2} - 2c)\sigma^2 B \frac{N - mp}{(N - mp + 2)(mp - p_1)} \\ & \sum_{k=1}^{m} \{tr\,[\underline{G}_k \underline{C}^{-1}_{k22.1}] + tr\,[(\underline{C}^{-1}_{k11} - \underline{C}^{-1}) \underline{X}^T_{k1} \underline{Q}_k \underline{X}_{k1}]\}. \end{aligned} \tag{3.67}$$

For $mp - p_1 \geq 3$, the above risk is minimized at $c = mp - p_1 - 2$.

Proof of Theorem 3.2.6. From the proof of Theorem 2.2.8, it can be shown that

$$\begin{aligned} & r(\xi, \underline{e}_{1B}, \cdots, \underline{e}_{mB}) \\ = \; & \sigma^2(1 - B) \sum_{k=1}^{m} tr\,[(\underline{X}_k^T \underline{V}_k^{-1} \underline{X}_k)^{-1} \underline{X}_k^T \underline{Q}_k \underline{X}_k]. \end{aligned} \tag{3.68}$$

Then, Eq. (1.66) can be proved by combining (1.58) and (1.67).

3.3 Empirical Bayes Subset Estimators

In this section four empirical Bayes subset estimators of $\underline{\beta}_{k1}$ are discussed and compared, under the loss

$$L(\underline{\beta}_{k1}; \underline{a}_k) = (\underline{a}_k - \underline{\beta}_{k1})^T \underline{U}_k (\underline{a}_k - \underline{\beta}_{k1}). \tag{3.69}$$

The corresponding shrinkage coefficient is derived from the usual F ratio for testing $H_0 : \underline{\beta}_{k2} = \underline{0}$. If $\underline{e}^1_{\phi k1}$ is defined as an estimator of $\underline{\beta}_{k1}$ for some k that shrinks the unrestricted least squares estimator $\tilde{\underline{\beta}}_{k1}$ to the restricted least squares estimator $\hat{\underline{\beta}}_{k1}$ under the hypothesis $H_0 : \underline{\beta}_{k2} = \underline{0}$, then $\underline{e}^1_{\phi k1}$ can be expressed as

$$\underline{e}^1_{\phi k1} = \tilde{\underline{\beta}}_{k1} - \frac{\phi(F_k)}{F_k} (\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{k1}), \tag{3.70}$$

where $F_k = (N - mp + 2) \tilde{\underline{\beta}}^T_{k2} \underline{C}_{k22.1} \tilde{\underline{\beta}}_{k2} / SSE$.

Now, $\underline{e}^2_{\phi k1}$ is defined as an estimator that shrinks the unrestricted least squares estimator $\tilde{\underline{\beta}}_{k1}$ to the common restricted least squares estimator $\hat{\underline{\beta}}_{1.}$ under the hypothesis $H_0 : \underline{\beta}_{k1} = \underline{\beta}_{1.}$ and $\underline{\beta}_{k2} = \underline{0}$ for all k. If the shrinkage coefficient, $\phi(F_k)/F_k$, is the same one described in $\underline{e}^1_{k1}$, then

$$\underline{e}^2_{\phi k1} = \tilde{\underline{\beta}}_{k1} - \frac{\phi(F_k)}{F_k}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.}). \tag{3.71}$$

Replacing the shrinkage coefficient $\phi(F_k)/F_k$ by $\phi(F)/F$ of $\underline{e}^1_{\phi k1}$,

$$\underline{e}^3_{\phi k1} = \tilde{\underline{\beta}}_{k1} - \frac{\phi(F)}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{k1}), \tag{3.72}$$

where $F = (N - mp + 2)(SSH_1 + SSH_2)/SSE$ has been mentioned in Section 3.2.

The last one is a subset of $\underline{e}_{k\phi}$ described in Section 3.2. Thus, $\underline{e}^4_{\phi k1}$ can be expressed as

$$\underline{e}^4_{\phi k1} = \tilde{\underline{\beta}}_{k1} - \frac{\phi(F)}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.}). \tag{3.73}$$

The comparison among those estimators are based on their frequentist risks and Bayes risks. Now, these estimators of $\underline{\beta}_{k1}$ are first evaluated for some individual k according to the Bayes criterion. Let $\phi(F) = \phi(F_k) = c$, a positive constant, and rename $\underline{e}^i_{\phi k1}$ as $\underline{e}^i_{ck1}$ for $i = 1, 2, 3, 4$. The following theorem shows that under the prior described in Section 3.2, the Bayes risk of $\underline{e}^4_{ck1}$ with $c = mp - p_1 - 2$, i.e., the subset of $\underline{e}_{kEB}$, is the smallest one among the four types estimators of $\underline{\beta}_{k1}$ irrespective of any p.d. $\underline{U}_k$.

Theorem 3.3.1. Consider the prior ξ in Theorem 3.2.4. Let $r(\xi, \underline{e}_{k1})$ denote the Bayes risk of an estimator $\underline{e}_{k1}$ of $\underline{\beta}_{k1}$ for some individual k under the prior ξ when the loss is given in (3.69). Then,

$$\begin{aligned} & \min_c\{r(\xi, \underline{e}^4_{ck1})\} \\ = \; & \min\{\min_c\{r(\xi, \underline{e}^1_{ck1})\}, \min_c\{r(\xi, \underline{e}^2_{ck1})\}, \\ & \min_c\{r(\xi, \underline{e}^3_{ck1})\}, \min_c\{r(\xi, \underline{e}^4_{ck1})\}\}, \end{aligned}$$

$$\begin{aligned} & \min_c\{r(\xi, \underline{e}^1_{ck1})\} \\ = \ & \max\{\min_c\{r(\xi, \underline{e}^1_{ck1})\}, \min_c\{(\xi, \underline{e}^2_{ck1})\}, \\ & \min_c\{r(\xi, \underline{e}^3_{ck1})\}, \min_c\{(\xi, \underline{e}^4_{ck1})\}\}. \end{aligned}$$

Proof of Theorem 3.3.1. By expressing $\underline{e}^1_{ck1} - \underline{\beta}_{k1}$ as

$$\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1} - \frac{c}{F_k}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{k1}), \tag{3.74}$$

one gets

$$\begin{aligned} r(\xi, \underline{e}^1_{ck1}) \ = \ & \sigma^2 tr\,[\,\underline{C}^{-1}_{k11.2}\underline{U}_k\,] \\ - \ & 2E\{\frac{c}{F_k}(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1})^T\underline{U}_k(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{k1})\} \\ + \ & E\{\frac{c^2}{F_k^2}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{k1})^T\underline{U}_k(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{k1})\}. \end{aligned} \tag{3.75}$$

Recall $B = \sigma^2/(\sigma^2 + \tau^2)$. From (3.6), (3.11) and the mutual independence among $\hat{\underline{\beta}}_{k1}$'s, $\tilde{\underline{\beta}}_{k2}$'s and SSE,

$$\begin{aligned} & E(\frac{c}{F_k}(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1})(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{k1})^T) \\ = \ & E(\frac{c}{F_k}(\tilde{\underline{\beta}}_{k1} - E(\underline{\beta}_{k1}|\underline{Y}_k))(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{k1})^T) \\ = \ & \sigma^2 cB\frac{N - mp}{N - mp + 2}E(\frac{\underline{C}^{-1}_{k11}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2}\tilde{\underline{\beta}}^T_{k2}\underline{C}_{k21}\underline{C}^{-1}_{k11}}{\tilde{\underline{\beta}}^T_{k2}\underline{C}_{k22.1}\tilde{\underline{\beta}}_{k2}}). \end{aligned} \tag{3.76}$$

Now, by applying Basu's theorem, Eq. (3.76) equals

$$\sigma^2 cB\frac{N - mp}{N - mp + 2}\frac{\underline{C}^{-1}_{k11.2} - \underline{C}^{-1}_{k11}}{p_2}. \tag{3.77}$$

Similarly,

$$\begin{aligned} & E(\frac{c^2}{F_k^2}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{k1})(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{k1})^T) \\ = \ & \sigma^4 c^2\frac{N - mp}{N - mp + 2}E(\frac{\underline{C}^{-1}_{k11}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2}\tilde{\underline{\beta}}^T_{k2}\underline{C}_{k21}\underline{C}^{-1}_{k11}}{(\tilde{\underline{\beta}}^T_{k2}\underline{C}_{k22.1}\tilde{\underline{\beta}}_{k2})^2}) \\ = \ & \sigma^2 c^2 B\frac{N - mp}{N - mp + 2}\frac{\underline{C}^{-1}_{k11.2} - \underline{C}^{-1}_{k11}}{(p_2)(p_2 - 2)}. \end{aligned} \tag{3.78}$$

By using (3.77) and (3.78), it can be then shown that

$$\begin{aligned} r(\xi, \underline{e}^1_{ck1}) &= \sigma^2 tr\,[\,\underline{C}^{-1}_{k11.2}\underline{U}_k\,] \\ &+ \sigma^2 B \frac{N-mp}{(N-mp+2)p_2}\Big(\frac{c^2}{p_2-2} - 2c\Big) \\ &\quad tr\,[\,(\underline{C}^{-1}_{k11.2} - \underline{C}^{-1}_{k11})\underline{U}_k\,], \end{aligned} \tag{3.79}$$

which is minimized at $c = p_2 - 2$. The minimum of $r(\xi, \underline{e}^1_{ck1})$ is

$$\begin{aligned} r(\xi, \underline{e}^1_{(p_2-2)k1}) &= \sigma^2 tr\,[\,\underline{C}^{-1}_{k11.2}\underline{U}_k\,] \\ &- \sigma^2 B \frac{(N-mp)(p_2-2)}{(N-mp+2)p_2} tr\,[\,(\underline{C}^{-1}_{k11.2} - \underline{C}^{-1}_{k11})\underline{U}_k\,]. \end{aligned} \tag{3.80}$$

By using the same arguments described above, the mutual independence of $\hat{\underline{\beta}}_{1.}$, $\tilde{\underline{\beta}}_{k2}$ and $\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.}$, and the fact that $E(\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})(\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})^T = V(\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})$ $= (\sigma^2 + \tau^2)(\underline{C}^{-1}_{k11} - \underline{C}^{-1})$, one gets

$$\begin{aligned} r(\xi, \underline{e}^2_{ck1}) &= \sigma^2 tr\,[\,\underline{C}^{-1}_{k11.2}\underline{U}_k\,] \\ &+ \sigma^2 B \frac{N-mp}{(N-mp+2)p_2}\Big(\frac{c^2}{p_2-2} - 2c\Big) \\ &\quad tr\,[\,(\underline{C}^{-1}_{k11.2} - \underline{C}^{-1})\underline{U}_k\,], \end{aligned} \tag{3.81}$$

which is minimized at $c = p_2 - 2$. Thus, the minimum of $r(\xi, \underline{e}^2_{ck1})$ can be expressed as

$$\begin{aligned} r(\xi, \underline{e}^2_{(p_2-2)k1}) &= \sigma^2 tr\,[\,\underline{C}^{-1}_{k11.2}\underline{U}_k\,] \\ &- \sigma^2 B \frac{(N-mp)(p_2-2)}{(N-mp+2)p_2} \\ &\times tr\,[\,(\underline{C}^{-1}_{k11.2} - \underline{C}^{-1})\underline{U}_k\,]. \end{aligned} \tag{3.82}$$

Recall that $\underline{C}_{k11} = \underline{X}^T_{k1}\underline{V}^{-1}_k\underline{X}_{k1}$ and $\underline{C} = \sum_{k=1}^m \underline{C}_{k11}$.

The Bayes risk of $\underline{e}^3_{ck1}$ also can be expressed as

$$\begin{aligned} r(\xi, \underline{e}^3_{ck1}) &= \sigma^2 tr\,[\,\underline{C}^{-1}_{k11.2}\underline{U}_k\,] \\ &+ \sigma^2 B \frac{N-mp}{(N-mp+2)(mp-p_1)}\Big(\frac{c^2}{mp-p_1-2} - 2c\Big) \\ &\quad tr\,[\,(\underline{C}^{-1}_{k11.2} - \underline{C}^{-1}_{k11})\underline{U}_k\,], \end{aligned} \tag{3.83}$$

which is minimized at $c = mp - p_1 - 2$. The minimum of $r(\xi, \underline{e}^3_{ck1})$ can be shown as

$$\begin{aligned} r(\xi, \underline{e}^2_{(mp-p_1-2)k1}) &= \sigma^2 tr\,[\,\underline{C}^{-1}_{k11.2}\underline{U}_k\,] \\ &- \sigma^2 B\frac{(N-mp)(mp-p_1-2)}{(N-mp+2)(mp-p_1)} \\ &\times \; tr\,[\,(\underline{C}^{-1}_{k11.2} - \underline{C}^{-1}_{k11})\underline{U}_k\,]. \end{aligned} \tag{3.84}$$

By using (3.54),

$$\begin{aligned} r(\xi, \underline{e}^4_{ck1}) &= \sigma^2 tr\,[\,\underline{C}^{-1}_{k11.2}\underline{U}_k\,] \\ &+ \sigma^2 B\frac{N-mp}{(N-mp+2)(mp-p_1)}\left(\frac{c^2}{mp-p_1-2} - 2c\right) \\ &\quad tr\,[\,(\underline{C}^{-1}_{k11.2} - \underline{C}^{-1})\underline{U}_k\,], \end{aligned} \tag{3.85}$$

which is minimized at $c = mp - p_2 - 1$. The minimum of $r(\xi, \underline{e}^4_{ck1})$ is given by

$$\begin{aligned} r(\xi, \underline{e}^2_{(mp-p_1-2)k1}) &= \sigma^2 tr\,[\,\underline{C}^{-1}_{k11.2}\underline{U}_k\,] \\ &- \sigma^2 B\frac{(N-mp)(mp-p_1-2)}{(N-mp+2)(mp-p_1)} \\ &\times \; tr\,[\,(\underline{C}^{-1}_{k11.2} - \underline{C}^{-1})\underline{U}_k\,]. \end{aligned} \tag{3.86}$$

Note that $\underline{C}^{-1}_{k11.2} - \underline{C}^{-1}_{k11} = \underline{C}^{-1}_{k11}\underline{C}_{k12}\underline{C}^{-1}_{k22.1}\underline{C}_{k21}\underline{C}^{-1}_{k11}$ and $\underline{C}^{-1}_{k11.2} - \underline{C}^{-1} = \underline{C}^{-1}_{k11}\underline{C}_{k12}\underline{C}^{-1}_{k22.1}\underline{C}_{k21}\underline{C}^{-1}_{k11} + \underline{C}^{-1}_{k11} - \underline{C}^{-1}$. Comparing (3.80),(3.82),(3.84) and (3.86), Theorem 3.3.1 follows.

The condition described in Theorem 3.3.1 is clearly expected. It is only for $\underline{e}^4_{ck1}$ that the estimators of $\underline{\nu}_1$ as well as $1/(\sigma^2 + \tau^2)$ use all the available data, i.e., $\underline{Y}_1, \cdots, \underline{Y}_m$. The other three estimators use only $\underline{Y}_k$ but not all the $\underline{Y}_1, \cdots, \underline{Y}_m$ for estimating either $\underline{\nu}_1$ or $1/(\sigma^2 + \tau^2)$ or both.

If c is assumed to satisfy the conditions of that $0 < c < 2(p_2 - 2)$ in $\underline{e}^1_{ck1}$ and $\underline{e}^2_{ck1}$, and $0 < c < 2(mp - p_1 - 2)$ in $\underline{e}^3_{ck1}$ and $\underline{e}^4_{ck1}$, it is then easy to demonstrate that all these four estimators of $\underline{\beta}_{k1}$ are better than $\tilde{\underline{\beta}}_{k1}$.

Suppose one expresses

$$\begin{aligned} \hat{\underline{\beta}}_{k1} - \underline{\beta}_{k1} &= (\hat{\underline{\beta}}_{k1} - E_{\underline{\beta}_1, \cdots, \underline{\beta}_m, \sigma^2}(\hat{\underline{\beta}}_{k1})) \\ &+ \underline{C}^{-1}_{k11}\underline{C}_{k12}\underline{\beta}_{k2}, \end{aligned}$$

and gets

$$\begin{aligned} & r(\xi, \hat{\underline{\beta}}_{k1}) \\ = \; & \sigma^2 tr(\underline{C}_{k11}^{-1}\underline{U}_k) + \tau^2 tr\,[\,(\underline{C}_{k11.2}^{-1} - \underline{C}_{k11}^{-1})\underline{U}_k). \end{aligned} \tag{3.87}$$

Therefore,

$$\begin{aligned} & r(\xi, \hat{\underline{\beta}}_{k1}) - r(\xi, \underline{e}^1_{(p_2-2)k1}) \\ = \; & (\tau^2 - \sigma^2 + \sigma^2 B\frac{(N-mp)(p_2-2)}{(N-mp+2)p_2})tr\,[\,(\underline{C}_{k11.2}^{-1} - \underline{C}_{k11}^{-1})\underline{U}_k\,] \\ > \; & 0 \\ & \text{if and only if } (1-B)^2/B^2 > 1 - \frac{(N-mp)(p_2-2)}{(N-mp+2)p_2}, \end{aligned} \tag{3.88}$$

and

$$\begin{aligned} & r(\xi, \hat{\underline{\beta}}_{k1}) - r(\xi, \underline{e}^2_{(p_2-2)k1}) \\ = \; & \sigma^2 B\frac{(N-mp)(p_2-2)}{(N-mp+2)p_2})tr\,[\,(\underline{C}_{k11}^{-1} - \underline{C}^{-1})\underline{U}_k\,] \\ + \; & (\tau^2 - \sigma^2 + \sigma^2 B\frac{(N-mp)(p_2-2)}{(N-mp+2)p_2})tr\,[\,(\underline{C}_{k11.2}^{-1} - \underline{C}_{k11}^{-1})\underline{U}_k\,] \\ > \; & 0, \\ & \text{if and only if } (1-B)^2/B^2 > 1 - \frac{(N-mp)(p_2-2)}{(N-mp+2)p_2}. \end{aligned} \tag{3.89}$$

Also based on the same procedure given above,

$$\begin{aligned} & r(\xi, \hat{\underline{\beta}}_{k1}) - r(\xi, \underline{e}^3_{(mp-p_1-2)k1}) \\ = \; & (\tau^2 - \sigma^2 + \sigma^2 B\frac{(N-mp)(mp-p_1-2)}{(N-mp+2)(mp-p_1)})tr[\,(\underline{C}_{k11.2}^{-1} - \underline{C}_{k11}^{-1})\underline{U}_k\,] \\ > \; & 0, \\ & \text{if and only if } (1-B)^2/B^2 > 1 - \frac{(N-mp)(mp-p_1-2)}{(N-mp+2)(mp-p_1)}, \end{aligned} \tag{3.90}$$

and

$$\begin{aligned} & r(\xi, \hat{\underline{\beta}}_{k1}) - r(\xi, \underline{e}^4_{(mp-p_1-2)k1}) \\ = \; & \sigma^2 B\frac{(N-mp)(mp-p_1-2)}{(N-mp+2)(mp-p_1)})tr\,[\,(\underline{C}_{k11}^{-1} - \underline{C}^{-1})\underline{U}_k\,] \end{aligned}$$

$$
\begin{aligned}
&+ \quad (\tau^2 - \sigma^2 + \sigma^2 B \frac{(N-mp)(mp-p_1-2)}{(N-mp+2)(mp-p_1)}) tr\,[\,(\underline{C}_{k11.2}^{-1} - \underline{C}_{k11}^{-1})\underline{U}_k\,] \\
&> \quad 0, \qquad\qquad (3.91) \\
&\qquad \text{if and only if } (1-B)^2/B^2 > 1 - \frac{(N-mp)(mp-p_1-2)}{(N-mp+2)(mp-p_1)}.
\end{aligned}
$$

The results of (3.88), (3.89), (3.90), and (3.91) are not surprising since $\hat{\underline{\beta}}_{k1}$ is expected to perform better than $\underline{e}_{ck1}^{i}$'s (i=1,2,3,4) when the prior variance τ^2 is much smaller than the sampling variance σ^2. On the other way, the sampling variance σ^2 is smaller than the prior variance τ^2 i.e., there is no strong evidence to show $\underline{\beta}_{k2} = \underline{0}$, then all these four $\underline{e}_{ck1}^{i}$'s can have smaller Byes risk than $\hat{\underline{\beta}}_{k1}$.

Next, the frequentist criterion is utilized to evaluate the estimators.

Recall (3.6) and the mutual independence among $\hat{\underline{\beta}}_{k1}$'s, $\tilde{\underline{\beta}}_{k2}$'s and SSE under the the distributions of $(\underline{Y}_k|\underline{\beta}_k, \sigma^2)$, $k = 1, \cdots, m$. The frequentist risk of $\underline{e}_{\phi k1}^{1}$ under the loss in (3.69) is

$$
\begin{aligned}
& R(\underline{\beta}_{k1}, \underline{e}_{\phi k1}^{1}) \\
= \;& E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2} tr\,[\,(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1})(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1})^T \underline{U}_k\,] \\
- \;& 2tr\,[\,\underline{C}_{k11}^{-1}\underline{C}_{k12}\, E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{(\tilde{\underline{\beta}}_{k2} - \underline{\beta}_{k2})\frac{\phi(F_k)}{F_k}(\underline{C}_{k11}^{-1}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2})^T\}\underline{U}_k\,] \\
+ \;& E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}(\frac{\phi^2(F_k)}{F_k^2}\tilde{\underline{\beta}}_{k2}^T \underline{C}_{k21}\underline{C}_{k11}^{-1}\underline{U}_k\underline{C}_{k11}^{-1}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2}). \qquad (3.92)
\end{aligned}
$$

As mentioned earlier, by using a generalization of Stein's identity (Berger and Haff, 1983),

$$
\begin{aligned}
& E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{(\tilde{\underline{\beta}}_{k2} - \underline{\beta}_{k2})\frac{\phi(F_k)}{F_k}(\underline{C}_{k11}^{-1}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2})^T\} \\
= \;& \sigma^2 \underline{C}_{k22.1}^{-1} E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}(\frac{\partial}{\partial \tilde{\underline{\beta}}_{k2}}\frac{\phi(F_k)}{F_k}\tilde{\underline{\beta}}_{k2}^T \underline{C}_{k21}\underline{C}_{k11}^{-1}) \\
= \;& \sigma^2 E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}(2(\frac{\phi'(F_k)}{F_k} - \frac{\phi(F_k)}{F_k^2}) \\
\times \;& \underline{C}_{k11}^{-1}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2}\tilde{\underline{\beta}}_{k2}^T \underline{C}_{k21}\underline{C}_{k11}^{-1}/(SSE/(N-mp+2)) \\
+ \;& \frac{\phi(F_k)}{F_k}(\underline{C}_{k11.2}^{-1} - \underline{C}_{k11}^{-1}). \qquad (3.93)
\end{aligned}
$$

Combining (3.92) and (3.93) yields

$$
\begin{aligned}
& R(\underline{\beta}_{k1}, \underline{e}^1_{\phi k1}) \\
= \; & \sigma^2 tr\,[\,\underline{C}^{-1}_{k11.2}\underline{U}_k\,] - 2\sigma^2 E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{2(\frac{\phi'(F_k)}{F_k} - \frac{\phi(F_k)}{F_k^2}) \\
\times \; & \tilde{\underline{\beta}}^T_{k2}\underline{C}_{k21}\underline{C}^{-1}_{k11}\underline{U}_k\underline{C}^{-1}_{k11}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2}/(SSE/(N-mp+2)) \\
+ \; & \frac{\phi(F_k)}{F_k} tr\,[\,(\underline{C}^{-1}_{k11.2} - \underline{C}^{-1}_{k11})\underline{U}_k\,]\} \\
+ \; & E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{(\frac{\phi^2(F_k)}{F_k^2}(\tilde{\underline{\beta}}^T_{k2}\underline{C}_{k21}\underline{C}^{-1}_{k11}\underline{U}_k\underline{C}^{-1}_{k11}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2})\}. \qquad (3.94)
\end{aligned}
$$

The frequentist risk of $\underline{e}^2_{\phi k1}$ can be then obtained as

$$
\begin{aligned}
& R(\underline{\beta}_{k1}, \underline{e}^2_{\phi k1}) \\
= \; & E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2} tr\,[\,(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1})(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1})^T\underline{U}_k\,] \\
- \; & 2tr\,[\,E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{((\hat{\underline{\beta}}_{k1} - E(\hat{\underline{\beta}}_{k1})) - \underline{C}^{-1}_{k11}\underline{C}_{k12}(\tilde{\underline{\beta}}_{k2} - \underline{\beta}_{k2})) \\
\times \; & \frac{\phi(F_k)}{F_k}(\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.} - \underline{C}^{-1}_{k11}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2})^T\}\underline{U}_k\,] \\
+ \; & E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{\frac{\phi^2(F_k)}{F_k^2}((\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})^T\underline{U}_k(\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.}) \\
- \; & 2(\underline{C}^{-1}_{k11}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2})^T\underline{U}_k(\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.}) + \tilde{\underline{\beta}}^T_{k2}\underline{C}_{k21}\underline{C}^{-1}_{k11}\underline{U}_k\underline{C}^{-1}_{k11}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2})\}. \qquad (3.95)
\end{aligned}
$$

By using a generalization of Stein's identity (Berger and Haff, 1983), the cross-product term of (3.95) can be expressed as

$$
\begin{aligned}
& 2\sigma^2 tr\,[\,E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{2(\frac{\phi'(F_k)}{F_k} - \frac{\phi(F_k)}{F_k^2})(\underline{C}^{-1}_{k11}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2}\tilde{\underline{\beta}}^T_{k2}\underline{C}_{k21}\underline{C}^{-1}_{k11} \\
- \; & \underline{C}^{-1}_{k11}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2}(\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})^T)/(SSE/(N-mp+2)) \\
+ \; & \frac{\phi(F_k)}{F_k}(\underline{C}^{-1}_{k11.2} - \underline{C}^{-1})\}\underline{U}_k\,]. \qquad (3.96)
\end{aligned}
$$

From (3.95) and (3.96), one gets

$$
\begin{aligned}
& R(\underline{\beta}_{k1}, \underline{e}^2_{\phi k1}) \\
= \; & \sigma^2 tr(\underline{C}^{-1}_{k11.2}\underline{U}_k) - 2\sigma^2 E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{2(\frac{\phi'(F_k)}{F_k} - \frac{\phi(F_k)}{F_k^2}) \\
\times \; & (\tilde{\underline{\beta}}^T_{k2}\underline{C}_{k21}\underline{C}^{-1}_{k11}\underline{U}_k\underline{C}^{-1}_{k11}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2} - (\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})^T\underline{U}_k\underline{C}^{-1}_{k11}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2})
\end{aligned}
$$

$$
\begin{aligned}
&\div \ (SSE/(N-mp+2)) \\
&+ \ \frac{\phi(F_k)}{F_k} tr\,[\,(\underline{C}_{k11.2}^{-1} - \underline{C}^{-1})\underline{U}_k\,]\} \\
&+ \ E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{\frac{\phi^2(F_k)}{F_k^2}((\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})^T\underline{U}_k(\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.}) \\
&- \ 2(\underline{C}_{k11}^{-1}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2})^T\underline{U}_k(\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.}) + \tilde{\underline{\beta}}_{k2}^T\underline{C}_{k21}\underline{C}_{k11}^{-1}\underline{U}_k\underline{C}_{k11}^{-1}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2})\}. \quad (3.97)
\end{aligned}
$$

By using the same argument, it can be shown that

$$
\begin{aligned}
&R(\underline{\beta}_{k1}, \underline{\varepsilon}_{\phi k1}^3) \\
&= \ \sigma^2 tr(\underline{C}_{k11.2}^{-1}\underline{U}_k) - 2\sigma^2 E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{2(\frac{\phi'(F)}{F} - \frac{\phi(F)}{F^2}) \\
&\times \ (\tilde{\underline{\beta}}_{k2}^T\underline{C}_{k21}\underline{C}_{k11}^{-1}\underline{U}_k\underline{C}_{k11}^{-1}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2} - (\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})^T\underline{U}_k\underline{C}_{k11}^{-1}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2}) \\
&\div \ (SSE/(N-mp+2)) \\
&+ \ \frac{\phi(F)}{F} tr((\underline{C}_{k11.2}^{-1} - \underline{C}_{k11}^{-1})\underline{U}_k)\} \\
&+ \ E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}(\frac{\phi^2(F)}{F^2}\tilde{\underline{\beta}}_{k2}^T\underline{C}_{k21}\underline{C}_{k11}^{-1}\underline{U}_k\underline{C}_{k11}^{-1}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2}), \qquad (3.98)
\end{aligned}
$$

and

$$
\begin{aligned}
&R(\underline{\beta}_{k1}, \underline{\varepsilon}_{\phi k1}^4) \\
&= \ \sigma^2 tr(\underline{C}_{k11.2}^{-1}\underline{U}_k) - 2\sigma^2 E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{2(\frac{\phi'(F)}{F} - \frac{\phi(F)}{F^2}) \\
&\quad ((\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})^T\underline{U}_k(\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.}) - 2(\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})^T\underline{U}_k\underline{C}_{k11}^{-1}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2} \\
&+ \ \tilde{\underline{\beta}}_{k2}^T\underline{C}_{k21}\underline{C}_{k11}^{-1}\underline{U}_k\underline{C}_{k11}^{-1}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2})/(SSE/(N-mp+2)) \\
&+ \ \frac{\phi(F)}{F} tr\,[\,(\underline{C}_{k11.2}^{-1} - \underline{C}^{-1})\underline{U}_k\,]\} \\
&+ \ E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{\frac{\phi^2(F)}{F^2}((\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})^T\underline{U}_k(\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.}) \\
&\quad -2(\hat{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{1.})^T\underline{U}_k\underline{C}_{k11}^{-1}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2} + \tilde{\underline{\beta}}_{k2}^T\underline{C}_{k21}\underline{C}_{k11}^{-1}\underline{U}_k\underline{C}_{k11}^{-1}\underline{C}_{k12}\tilde{\underline{\beta}}_{k2})\}. \quad (3.99)
\end{aligned}
$$

Investigations of Eqs.(3.94), (3.97), (3.98) and (3.99) indicate that there is no winner among these four estimators as demonstrated by the comparisons of their frequentist risk with the loss defined as (3.69). This remains true even though $\phi(F) = \phi(F_k)$ is equal to a positive constant c.

3.4 Numerical Examples

To illustrate the results for the estimation derived in the previous sections, a Monte-Carlo experiment was conducted to assess the frequentist risks as well as the Bayes risks of the different estimators.

Random samples of $m = 3$ normally distributed random vectors $\underline{Y}_k$ $(k = 1, 2, 3)$ of dimensions $n_1 = 16$, $n_2 = 14, n_3 = 17$ are taken in a 2^3 factorial experiment. Here $\underline{Y}_1$ is the vector responses of the $1st$ individual, which is based on a 2^3 factorial experiment replicated two times (see Section 2.4) and $\underline{Y}_i$'s $(i = 2, 3)$ are the ith individual's responses vector which are based on 2^3 factorial experiment with unequal replications for certain treatment combinations. Let Y_{ijk} denote the response of the kth individual receiving the ith treatment within the jth replicate. It is assumed that

$$\begin{aligned} E(Y_{ijk}) &= \gamma_{0k} + \gamma_{1k}\varsigma_{1i} + \gamma_{2k}\varsigma_{2i} + \gamma_{3k}\varsigma_{3i} \\ &+ \gamma_{12k}\varsigma_{1i}\varsigma_{2i} + \gamma_{13k}\varsigma_{1i}\varsigma_{3i} + \gamma_{23k}\varsigma_{2i}\varsigma_{3i} + \gamma_{123k}\varsigma_{1i}\varsigma_{2i}\varsigma_{3i}, \end{aligned}$$

where $i = 1, \cdots, 8$, $k = 1, 2, 3$. Note that in most cases, there are two responses corresponding to the kth individual receiving the ith treatment. The only exceptions are $(i, k) = (1, 2)$, $(6, 2)$ and $(2, 3)$. in the first two cases, there is one reponse. For the third case, there are three responses. Here $\varsigma_{1i} = 1$ for $i = 1, 2, 3, 4$, corresponds to the high level of ς_1, while $\varsigma_{1i} = -1$ for $i = 5, 6, 7, 8$, corresponds to the low level of ς_1. Also $\varsigma_{2i} = 1$ $(\varsigma_{3i} = 1)$ for $i = 1, 2, 5, 6$ $(i = 1, 3, 5, 7)$, corresponds to the high level of ς_2 (ς_3), and $\varsigma_{2i} = -1$ $(\varsigma_{3i} = -1)$ for $i = 3, 4, 7, 8$ $(i = 2, 4, 6, 8)$, corresponds to the low level of ς_2 (ς_3). For $\underline{\beta}_{k1}^T = (\gamma_{0k}, \gamma_{1k}, \gamma_{2k}, \gamma_{3k})$ and $\underline{\beta}_{k2}^T = (\gamma_{12k}, \gamma_{13k}, \gamma_{23k}, \gamma_{123k})$, $mp_1 = mp_2 = 12$ and $N = 16 + 14 + 17 = 47$. The design matrices $\underline{X}_k$'s can be partitioned as $\underline{X}_k = (\underline{X}_{k1}, \underline{X}_{k2})$, $k = 1, 2, 3$, where

$$\underline{X}_{11}^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \end{bmatrix},$$

$$\underline{X}_{12}^T = \begin{bmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \end{bmatrix},$$

$$\underline{X}_{21}^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & 1 & 1 & -1 & -1 \end{bmatrix},$$

$$\underline{X}_{22}^T = \begin{bmatrix} 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 \end{bmatrix},$$

and

$$\underline{X}_{31}^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \end{bmatrix},$$

$$\underline{X}_{32}^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \end{bmatrix},$$

It is assumed that $\underline{V}(\underline{Y}_1) = \underline{I}_{16}$, $\underline{V}(\underline{Y}_2) = \underline{I}_{14}$, $\underline{V}(\underline{Y}_3) = \underline{I}_{17}$, and $\underline{Q}_k = \underline{V}(\underline{Y}_k)$, for $k = 1, 2, 3$.

First the simulated frequentist risks are evaluated based on the loss as defined in (3.4). For a given set of $\underline{\beta}_k$ $(k = 1, 2, 3)$, random vectors $\underline{Y}_k$'s were generated from $N_{nk}(\underline{X}_k \underline{\beta}_k, \underline{I}_k)$ in each replication. The sampling procedure was repeated 1,000 times. The estimated frequentist risks of estimating $\underline{X}_k \underline{\beta}_k$ $(k = 1, 2, 3)$, i.e., the average squared error losses of the least squares estimators (LSE), the model in (3.1), $RLSE1$, the model in (3.3), $RLSE2$, the model in (3.2), the empirical Bayes estimator $(\underline{e}_{EB}^*)$, the positive part empirical Bayes estimator $(\underline{e}_{EB}^{+*})$, the preliminary test estimators $(\underline{e}_{PTE\alpha}^*$, α being the level of significance, =.10, .05, .01), the modified empirical Bayes estimators $(\underline{e}_{MEB\alpha}^*)$, and the postive part modified empirical Bayes estimators $(\underline{e}_{MEB\alpha}^{+*})$, are compared. The result are shown in Table 3.1.

Table 3.1: The estimated frequentist risk results for 2^3 factorial experiment.

The estimator of $\underline{X}_k\underline{\beta}_k$'s	$\underline{\beta}_1$	(24630000)	(24630000)	(24630000)	(24637129)
	$\underline{\beta}_2$	(24630000)	(25520000)	(24636 −1 25)	(24636 −1 25)
	$\underline{\beta}_3$	(24630000)	(36520000)	(24630000)	(24631111)
$\underline{e}^*_{LSE}$		23.9434	23.9434	23.9434	23.9434
$\underline{e}^*_{RLSE1}$		4.18885	69.3494	885.765	3112.11
$\underline{e}^*_{RLSE2}$		12.1528	12.1528	854.819	3082.02
$\underline{e}^*_{EB}$		7.54115	20.4126	23.6275	23.8590
$\underline{e}^{+*}_{EB}$		5.28671	20.4126	23.6275	23.8590
$\underline{e}^*_{PTE.10}$		6.59654	24.3170	23.9434	23.9434
$\underline{e}^*_{MEB.10}$		5.09850	20.7060	23.6275	23.8590
$\underline{e}^{+*}_{MEB.10}$		5.09850	20.7060	23.6275	23.8950
$\underline{e}^*_{PTE.05}$		5.56069	24.9046	23.9434	23.9434
$\underline{e}^*_{MEB.05}$		4.78082	21.2411	23.6275	23.8590
$\underline{e}^{+*}_{MEB.05}$		4.78082	21.2411	23.6275	23.8590
$\underline{e}^*_{PTE.01}$		4.71660	28.9008	23.9434	23.9434
$\underline{e}^*_{MEB.01}$		4.33754	25.1467	23.6275	23.8590
$\underline{e}^{+*}_{MEB.01}$		4.33754	25.1467	23.6275	23.8590

* Note: For the right two columns, during the 1,000 replications, the null hypothesis, $\underline{\beta}_{k1} = \underline{\beta}_{1.}$ and $\underline{\beta}_{k2} = \underline{0}$ $(k = 1, 2, 3)$, was rejected in every replication for all α =.10, .05 and .01. Hence, $\underline{e}_{kPTE\alpha} = \underline{X}_k\tilde{\underline{\beta}}_k$, $\underline{e}_{kMEB\alpha} = \underline{e}_{kEB}$ and $\underline{e}^+_{kMEB\alpha} = \underline{e}^+_{kEB}$.

When $\underline{\beta}_{11} = \underline{\beta}_{21} = \underline{\beta}_{31}$ and $\underline{\beta}_{k2} = \underline{0}$ $(k = 1, 2, 3)$, $\underline{\epsilon}^*_{RLSE1} = ((\underline{X}_{11}\hat{\underline{\beta}}_{1.})^T, (\underline{X}_{21}\hat{\underline{\beta}}_{1.})^T, (\underline{X}_{31}\hat{\underline{\beta}}_{1.})^T)^T$ is the best estimator of $(\underline{X}\underline{\beta})^* = ((\underline{X}_1\underline{\beta}_1)^T, (\underline{X}_2\underline{\beta}_2)^T, (\underline{X}_3\underline{\beta}_3)^T)^T$. However, in case $\underline{\beta}_{k2} = \underline{0}$ $(k = 1, 2, 3)$, then $\underline{\epsilon}^*_{RLSE2}$=$((\underline{X}_{11}\hat{\underline{\beta}}_{11})^T, (\underline{X}_{21}\hat{\underline{\beta}}_{21})^T, (\underline{X}_{31}\hat{\underline{\beta}}_{31})^T)^T$ is the best as an estimator of $(\underline{X}\underline{\beta})^*$. Substantial improvement in the estimated frequentist risk of $\underline{\epsilon}^{+*}_{EB}$ over $\underline{\epsilon}^*_{LSE} = ((\underline{X}_1\tilde{\underline{\beta}}_1)^T, (\underline{X}_2\tilde{\underline{\beta}}_2)^T, (\underline{X}_3\tilde{\underline{\beta}}_3)^T)^T$, the LSE of the model in (3.1), was observed in the cases of $\underline{\beta}_{k2} = \underline{0}$ $(k = 1, 2, 3)$. When $\underline{\beta}_{k1}$'s are all identical, the estimated frequentist risk of $\underline{\epsilon}^{+*}_{EB}$ is found only about one forth of magnitude as that of $\underline{\epsilon}^*_{LSE}$. As proved in the earlier section, the estimators $\underline{\epsilon}^*_{MEB}$ or $\underline{\epsilon}^{+*}_{MEB}$ performed better than $\underline{\epsilon}^*_{PTE}$. Finally, if $\underline{\beta}_{k2} \neq \underline{0}$ for some k, $\underline{\epsilon}^*_{EB}$ or $\underline{\epsilon}^{+*}_{EB}$ has the smallest frequentist risks. Admittedly, the risk improvement is not significant over $\underline{\epsilon}^*_{LSE}$. All other estimators are superior over $\underline{\epsilon}^*_{RLSE1}$ and $\underline{\epsilon}^*_{RLSE2}$ with the considerably less estimated frequentist risks.

Table 3.2 shows the estimated Bayes risks of all the estimators. First, $\underline{\beta}_k$ $(k = 1, 2, 3)$ were generated seperately from $N(\underline{\nu}, \tau^2(\underline{X}_k^T\underline{X}_k)^{-1})$, where $\underline{\nu}$ and τ^2 are the given-values. For a generated set of $\underline{\beta}_k$'s and a given σ^2, random vectors $\underline{Y}_k$ $(k = 1, 2, 3)$ were then genterated seperately from $N(\underline{X}_k\underline{\beta}_k, \sigma^2\underline{I}_{nk})$. Recall that $n_1 = 16, n_2 = 14, n_3 = 17$. Once again, the sampling procedure was repeated 1,000 times.

When $\tau^2 < \sigma^2$, i.e., when the prior variability is smaller than the sampling variability, the estimated risks of the different estimators maintain in almost the same order as the corresponding estimated frequentist risks. The case $\underline{\nu}_2 = \underline{0}$ $(\underline{\nu}_2 \neq \underline{0})$ corresponds to the case $\underline{\beta}_{11} = \underline{\beta}_{21} = \underline{\beta}_{31}$ and $\underline{\beta}_{k2} = \underline{0}$ for all k (for at least one of $\underline{\beta}_{k2} \neq \underline{0}$). The only difference is that $\underline{\epsilon}^{+*}_{EB}$ has the smallest estimated Bayes risk when $\underline{\nu}_2 = \underline{0}$. However, the relation between the Bayes risks of $\underline{\epsilon}^*_{EB}$ and $\underline{\epsilon}^*_{RLSE1}$, as dependent upon the relation of τ^2 and σ^2, when $\underline{\nu}_2 = \underline{0}$ agrees well with Theorem 3.2.5. As $\tau^2 > \sigma^2$, even though $\underline{\nu}_2 = \underline{0}$, $\underline{\epsilon}^*_{RLSE1}$ was seen to perform poorly as compared with others. The increased prior variability diminishes the importance of the prior mean. As $\tau^2 = \sigma^2$, then the $\underline{\epsilon}^*_{EB}$ and $\underline{\epsilon}^{+*}_{EB}$ dominate both the LSE's of the

Table 3.2: The estimated Bayes risk results for 2^3 factorial experiment.

The estimator of $\underline{X}_k\underline{\beta}_k$'s	$\underline{\nu}$	(3 5 1 6 0 0 0 0)			(3 5 1 6 4 0 2 5)		
	σ^2	4			4		
	τ^2	1	4	9	1	4	9
$\underline{e}^*_{LSE}$		95.2410	95.2410	95.2410	95.2410	95.2410	95.2410
$\underline{e}^*_{RLSE1}$		36.1561	95.9029	195.481	2069.51	2129.74	2229.81
$\underline{e}^*_{RLSE2}$		60.2533	95.8641	155.215	2064.76	2100.34	2159.67
$\underline{e}^*_{EB}$		43.1769	62.7633	75.2881	93.0472	93.0843	93.1498
$\underline{e}^{+*}_{EB}$		33.2123	59.5154	74.6257	93.0472	93.0843	93.1498
$\underline{e}^*_{PTE.10}$		53.2020	95.1617	100.376	95.2410	95.2410	95.2410
$\underline{e}^*_{MEB.10}$		40.2773	72.3558	80.3389	93.0472	93.0843	93.1498
$\underline{e}^{+*}_{MEB.10}$		40.2773	72.3558	80.3389	93.0472	93.0843	93.1498
$\underline{e}^*_{PTE.05}$		46.3677	95.2423	104.812	95.2410	95.2410	95.2410
$\underline{e}^*_{MEB.05}$		39.2176	77.8064	85.5323	93.0472	93.0853	93.1498
$\underline{e}^{+*}_{MEB.05}$		39.2176	77.8064	85.5323	93.0472	93.0843	93.1498
$\underline{e}^*_{PTE.01}$		38.6651	95.4421	124.036	95.2410	95.2410	95.2410
$\underline{e}^*_{MEB.01}$		37.2853	87.1063	109.696	93.0472	93.0843	93.1498
$\underline{e}^{+*}_{MEB.01}$		37.2853	87.1063	109.696	93.0472	93.0843	93.1498

* Note: For $\underline{\nu} = (3\,5\,1\,6\,4\,0\,2\,5)$, during the 1,000 replications, the null hypothesis, $\underline{\beta}_{k1} = \underline{\beta}_1$ and $\underline{\beta}_{k2} = \underline{0}$ $(k = 1, 2, 3)$, was rejected in every replication for all $\alpha = .10, .05$ and $.01$. Hence, $\underline{e}_{kPTE\alpha} = \underline{X}_k\underline{\tilde{\beta}}_k$, $\underline{e}_{kMEB\alpha} = \underline{e}_{kEB}$ and $\underline{e}^+_{kMEB\alpha} = \underline{e}^+_{kEB}$.

model in (3.1) and the model in (3.3). The domination is more heavily pronounced in the case when $\nu_2 = \underline{0}$.

CHAPTER 4
EMPIRICAL BAYES ESTIMATION OF MULTIVARIATE REGRESSION MODEL II

4.1 Introduction

To estimate a multivariate linear regression model for m distinct individuals, there may be a choice between the following two models:

$$\underline{Y}_k = \underline{X}_{k1}\underline{\beta}_{k1} + \underline{X}_{k2}\underline{\beta}_{k2} + \underline{\varepsilon}_k, \tag{4.1}$$

and

$$\underline{Y}_k = \underline{X}_{k1}\underline{\beta}_{k1} + \underline{\varepsilon}_k, \tag{4.2}$$

where $\underline{Y}_k$'s, $\underline{X}_i$'s ($i = 1,2$), $\underline{\beta}_{k1}$'s, $\underline{\beta}_{k2}$'s, $\underline{\varepsilon}_k$'s, and $\underline{V}_k$'s are the same as defined in Section 3.1. Estimators of $\underline{X}_k\underline{\beta}_k$'s, once determined can be compared on the basis of general quadratic loss,

$$L(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m; \underline{a}_1, \cdots, \underline{a}_m) = \sum_{k=1}^{m} (\underline{a}_k - \underline{X}_k\underline{\beta}_k)^T \underline{Q}_k (\underline{a}_k - \underline{X}_k\underline{\beta}_k), \tag{4.3}$$

where $\underline{Q}_k$'s are some known p.d. matrices.

In this chapter, it is assumed that in (4.1) and (4.2), $\underline{X}_k = \underline{X}$ and $\underline{V}_k = \underline{V}$ $k = 1, \cdots, m$. The g-prior distribution of the coefficients $\underline{\beta}_k$'s is assumed here as

$$\begin{aligned} &\underline{\beta}_k \sim^{ind} N(\underline{\nu}_k, \underline{\Sigma}), k = 1, \cdots, m, \\ &\text{with } \underline{\nu}_k = \underline{D}\underline{a}_k, \ \underline{D} = \begin{pmatrix} \underline{D}_1 \\ \underline{0} \end{pmatrix}, \text{ and } \underline{\Sigma} = \tau^2(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}, \end{aligned} \tag{4.4}$$

where $\underline{a}_k$ represents the values of the concomitant 'background' regressor variables associated with the kth individual. If $\underline{A}^T = (\underline{a}_1, \cdots, \underline{a}_m)$ is an $r \times m$ 'across individual' design matrix of constants of full rank $r < m$, then $\underline{D}_1$ is a $p_1 \times r$ matrix

of unknown parameters. It should be noted that $\underline{D}$ is a $p \times r$ matrix and $p > p_1$. The structure of the prior mean displayed in (4.1) is modified from the one given in Reinsel (1985).

The empirical Bayes (EB) estimators of $(\underline{X\beta})^* = ((\underline{X\beta}_{\cdot 1})^T, \cdots, (\underline{X\beta}_{\cdot m})^T)^T$ will be introduced according to the above prior in Section 4.2. The EB estimators of $\underline{X\beta}_{\cdot k}$ serve as a weighted average of the LSE of $\underline{X\beta}_{\cdot k}$ based on the model in (4.1) and the weighted average of LSE's for the reduced model (4.2) for $k = 1, \cdots, m$. If a F ratio is defined from the hypothesis that $H_0 : \underline{\beta}_{k1} = \underline{D}_1 \underline{a}_k$ and $\underline{\beta}_{k2} = \underline{0}$, $k = 1, \cdots, m$, the empirical Bayes estimator of $\underline{X\beta}_{\cdot k}$ leans more towards $\underline{X\tilde{\beta}}_{\cdot k}$ for all k if the observed F ratio is large, and to the weighted average of the reduced LSE's, $\underline{X}_1\underline{\hat{\beta}}_{\cdot 11}, \cdots, \underline{X}_1\underline{\hat{\beta}}_{\cdot m1}$, if the observed F ratio is small.

A popular method of achieving the compromise estimator is the use of preliminary test estimators (PTE) based on the rejection or acceptance of the null hypothesis H_0. It is shown in Section 4.2 that for every PTE of $(\underline{X\beta})^*$, there is a corresponding modified EB estimator which dominates the PTE. This is not a surprising consequence, since the PTE does not take into account the degree of evidence for or against the null hypothesis $H_0 : \underline{\beta}_{k1} = \underline{D}_1 \underline{a}_k$ and $\underline{\beta}_{k2} = \underline{0}$ for all k.

Section 4.2 also introduces the positive part versions of the EB estimator. Several theorems are given, which describe the frequentist as well as the Bayesian properties of the proposed estimators. Section 4.3 introduces several estimators of $\underline{\beta}_{\cdot k1}$. These estimators are proposed in such ways that they shrink the unrestricted LSE, $\underline{\tilde{\beta}}_{\cdot k1}$, to the restricted LSE, $\underline{\hat{\beta}}_{\cdot k1}$, or a certain weighted average of $\underline{\hat{\beta}}_{\cdot k1}$'s. Under the squared error loss, their frequentist risks and Bayes risks are compared. It is shown that there are EB estimators of $(\underline{X\beta})^*$ which always dominate $\underline{e}^*_{LSE} = ((\underline{X\tilde{\beta}}_{\cdot 1})^T, \cdots, (\underline{X\tilde{\beta}}_{\cdot m})^T)^T$ under squared error loss. Finally, in Section 4.4, as a multivariate regression example, different estimators of $(\underline{X\beta})^*$ are compared according to their simulated frequentist and Bayes risks.

4.2 The Empirical Bayes Approach

In order to develop an EB procedure, a Bayes estimator of $\underline{X}\underline{\beta}_k$ for $k = 1, \cdots, m$, should first be obtained. Based on the prior distribution in (4.4), the posterior of $\underline{\beta}_k$ given $\underline{Y}_k$ is

$$N_p\left(\begin{pmatrix} \underline{D}_1\underline{a}_k \\ \underline{0} \end{pmatrix} + (1-B)\left(\tilde{\underline{\beta}}_k - \begin{pmatrix} \underline{D}_1\underline{a}_k \\ \underline{0} \end{pmatrix}\right), \sigma^2(1-B)(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\right), k = 1, \cdots, m, \tag{4.5}$$

where $\tilde{\underline{\beta}}_k = (\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\underline{X}^T\underline{V}^{-1}\underline{Y}_k$ is the $BLUE$ of $\underline{\beta}_k$, and $B = \sigma^2/(\sigma^2 + \tau^2)$. By expressing $\tilde{\underline{\beta}}_k^T = (\tilde{\underline{\beta}}_{k1}^T, \tilde{\underline{\beta}}_{k2}^T)$, the Bayes estimator of $\underline{X}\underline{\beta}_k$ under any quadratic loss (4.3) is

$$\begin{aligned} \underline{e}_{kB}(\underline{Y}_k) &= E(\underline{X}\underline{\beta}_k|\underline{Y}_k) = \underline{X}_1[\,\underline{D}_1\underline{a}_k + (1-B)(\tilde{\underline{\beta}}_{k1} - \underline{D}_1\underline{a}_k)\,] \\ &+ \quad (1-B)\underline{X}_2\tilde{\underline{\beta}}_{k2}, \text{ for } k = 1, \cdots, m. \end{aligned} \tag{4.6}$$

In an empirical Bayes framework, the parameters $\underline{D}_1$, σ^2 and τ^2 are unknown, and need to be estimated from the marginal distribution of $\underline{Y}_1, \cdots, \underline{Y}_m$. Note that marginally

$$\underline{Y}_k \sim^{ind.} N_n(\underline{X}_1\underline{D}_1\underline{a}_k, \sigma^2\underline{V} + \tau^2\underline{P}_{\underline{X}}), \tag{4.7}$$

where $\underline{P}_{\underline{X}} = \underline{X}(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\underline{X}^T$. By using Exercise 2.9, p.33 of Rao (1973), and recalling $B = \sigma^2/(\sigma^2 + \tau^2)$, it can be shown that

$$\begin{aligned} &(\sigma^2\underline{V} + \tau^2\underline{P}_{\underline{X}})^{-1} \\ &= \sigma^{-2}[\,\underline{V}^{-1} - (1-B)\underline{V}^{-1}\underline{P}_{\underline{X}}\underline{V}^{-1}\,]. \end{aligned} \tag{4.8}$$

Now, let $\underline{C}_{ij} = \underline{X}_i^T\underline{V}^{-1}\underline{X}_j$ $\underline{C}_{ii.j} = \underline{C}_{ii} - \underline{C}_{ij}\underline{C}_{jj}^{-1}\underline{C}_{ji}$ $(i, j = 1.2)$. Using (4.7) and (4.8), the maximum likelihood estimator of $\underline{D}_1$ is given by

$$\begin{aligned} \hat{\underline{D}}_1 &= (\underline{X}_1^T\underline{V}^{-1}\underline{X}_1)^{-1}\underline{X}_1^T\underline{V}^{-1}\underline{Y}^{\#T}\underline{A}(\underline{A}^T\underline{A})^{-1} \\ &= \hat{\underline{\beta}}_1^{\#T}\underline{A}(\underline{A}^T\underline{A})^{-1}, \end{aligned} \tag{4.9}$$

where $\underline{Y}^{\#} = (\underline{Y}_1, \cdots, \underline{Y}_m)^T$ and $\hat{\underline{\beta}}_1^{\#} = (\hat{\underline{\beta}}_{11}, \cdots, \hat{\underline{\beta}}_{m1})^T$, as $\hat{\underline{\beta}}_{k1} = (\underline{X}_1^T \underline{V}^{-1} \underline{X}_1)^{-1} \underline{X}_1 \underline{V}^{-1} \underline{Y}_k$. Recall $\underline{A}^T = (\underline{a}_1, \cdots, \underline{a}_m)$.

Then, based on the normal equation

$$\underline{X}_1 \underline{Y}_k = \underline{C}_{11} \hat{\underline{\beta}}_{k1} = \underline{C}_{11} \tilde{\underline{\beta}}_{k1} + \underline{C}_{12} \tilde{\underline{\beta}}_{k2}, \tag{4.10}$$

the following identity can be obtained

$$\tilde{\underline{\beta}}_{k1} = \hat{\underline{\beta}}_{k1} - \underline{C}_{11}^{-1} \underline{C}_{12} \tilde{\underline{\beta}}_{k2}. \tag{4.11}$$

Now, consider $\underline{Y}_k - \underline{X}_1 \underline{D}_1 \underline{a}_k = \underline{X} \tilde{\underline{\beta}}_k - \underline{X}_1 \underline{D}_1 \underline{a}_k + \underline{Y}_k - \underline{X} \tilde{\underline{\beta}}_k$. Since $\underline{P}_{\underline{X}} \underline{V}^{-1} \underline{X}_i = \underline{X}_i$ $(i = 1, 2)$, by using the idempotency of $\underline{V}^{-1} \underline{P}_{\underline{X}}$, and (4.8),

$$\begin{aligned}
& (\underline{Y}_k - \underline{X}_1 \underline{D}_1 \underline{a}_k)^T (\sigma^2 \underline{V} + \tau^2 \underline{P}_{\underline{X}})^{-1} (\underline{Y}_k - \underline{X}_1 \underline{D}_1 \underline{a}_k) \\
= \; & \sigma^{-2} \{ B [(\hat{\underline{\beta}}_{k1} - \underline{D}_1 \underline{a}_k)^T \underline{C}_{11} (\hat{\underline{\beta}}_{k1} - \underline{D}_1 \underline{a}_k) + \tilde{\underline{\beta}}_{k2}^T \underline{C}_{22.1} \tilde{\underline{\beta}}_{k2}] \\
+ \; & (\underline{Y}_k - \underline{X} \tilde{\underline{\beta}}_k)^T \underline{V}^{-1} (\underline{Y}_k - \underline{X} \tilde{\underline{\beta}}_k) \}.
\end{aligned}$$

Thus, adding all the k components of the above equation, and using the identity (4.11), one gets

$$\begin{aligned}
& \sum_{k=1}^{m} (\underline{Y}_k - \underline{X}_1 \underline{D}_1 \underline{a}_k)^T (\sigma^2 \underline{V} + \tau^2 \underline{P}_{\underline{X}})^{-1} (\underline{Y}_k - \underline{X}_1 \underline{D}_1 \underline{a}_k) \\
= \; & \sigma^{-2} \{ SSE + B \sum_{k=1}^{m} [(\hat{\underline{\beta}}_{k1} - \underline{D}_1 \underline{a}_k)^T \underline{C}_{11} (\hat{\underline{\beta}}_{k1} - \underline{D}_1 \underline{a}_k) \\
+ \; & \sum_{k=1}^{m} \tilde{\underline{\beta}}_{k2}^T \underline{C}_{22.1} \tilde{\underline{\beta}}_{k2}] \},
\end{aligned} \tag{4.12}$$

where

$$\begin{aligned}
& \sum_{k=1}^{m} (\hat{\underline{\beta}}_{k1} - \underline{D}_1 \underline{a}_k)^T \underline{C}_{11} (\hat{\underline{\beta}}_{k1} - \underline{D}_1 \underline{a}_k) \\
= \; & \sum_{k=1}^{m} \{ \underline{a}_k^T (\underline{D}_1^T \underline{C}_{11} \underline{D}_1) \underline{a}_k - 2 \underline{a}_k^T \underline{D}_1^T \underline{C}_{11} \hat{\underline{\beta}}_{k1} + \hat{\underline{\beta}}_{k1}^T \underline{C}_{11} \hat{\underline{\beta}}_{k1} \} \\
= \; & tr [(\underline{D}_1^T \underline{C}_{11} \underline{D}_1) \sum_{k=1}^{m} \underline{a}_k \underline{a}_k^T] - 2 tr [\underline{D}_1^T \underline{C}_{11} \sum_{k=1}^{m} \hat{\underline{\beta}}_{k1} \underline{a}_k^T] \\
+ \; & tr [\underline{C}_{11} \sum_{k=1}^{m} \hat{\underline{\beta}}_{k1} \hat{\underline{\beta}}_{k1}^T]
\end{aligned}$$

$$
\begin{aligned}
&= tr\,[\,(\underline{D}_1^T\underline{C}_{11}\underline{D}_1)\underline{A}^T\underline{A} - 2\underline{D}_1^T\underline{C}_{11}\hat{\underline{\beta}}_1^{\#T}\underline{A}\,] \\
&+ tr\,[\,\underline{C}_{11}\hat{\underline{\beta}}_1^{\#T}\hat{\underline{\beta}}_1^{\#}\,] \\
&= tr\{(\underline{D}_1 - \hat{\underline{\beta}}_1^{\#T}\underline{A}(\underline{A}^T\underline{A})^{-1})^T\underline{C}_{11}(\underline{D}_1 - \hat{\underline{\beta}}_1^{\#T}\underline{A}(\underline{A}^T\underline{A})^{-1})\underline{A}^T\underline{A}\} \\
&+ tr\,[\,(\hat{\underline{\beta}}_1^{\#T}\hat{\underline{\beta}}_1^{\#} - \hat{\underline{\beta}}_1^{\#T}\underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T\hat{\underline{\beta}}_1^{\#})\underline{C}_{11}\,]. \qquad (4.13)
\end{aligned}
$$

By using (4.9), the last term in (4.13) can be expressed as

$$
\begin{aligned}
& tr\,[\,\hat{\underline{\beta}}_1^{\#T}(\underline{I} - \underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A})\hat{\underline{\beta}}_1^{\#}\underline{C}_{11}\,] \qquad (4.14) \\
&= tr\,[\,(\hat{\underline{\beta}}_1^{\#T} - \hat{\underline{\beta}}_1^{\#T}\underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T)(\hat{\underline{\beta}}_1^{\#} - \underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T\hat{\underline{\beta}}_1^{\#})\underline{C}_{11}\,] \\
&= tr\,[\,(\hat{\underline{\beta}}_1^{\#T} - \hat{\underline{D}}_1\underline{A}^T)(\hat{\underline{\beta}}_1^{\#} - \underline{A}\hat{\underline{D}}_1^T)\underline{C}_{11}\,] \\
&= \sum_{k=1}^{m}(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T\underline{C}_{11}(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k). \qquad (4.15)
\end{aligned}
$$

Note that $SSE = \sum_{k=1}^{m}(\underline{Y}_k - \underline{X}\tilde{\underline{\beta}}_k)^T\underline{V}^{-1}(\underline{Y}_k - \underline{X}\tilde{\underline{\beta}}_k)$, the usual error sum of squares. Let $SSH_1 = \sum_{k=1}^{m}(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T\,\underline{C}_{11}(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)$, and $SSH_2 = \sum_{k=1}^{m}\tilde{\underline{\beta}}_{k2}^T\underline{C}_{22.1}\tilde{\underline{\beta}}_{k2}$. The summation of SSH_1 and SSH_2 is the sum of squares due to these hypotheses $H_{01} : \underline{\beta}_{k1} = \underline{D}_1\underline{a}_k$ for all k and $H_{02} : \underline{\beta}_{k2} = \underline{0}$ for all k. From (4.9), (4.12), (4.13), and (4.15), it is easy to see that

$(\hat{\underline{D}}_1, SSH_1 + SSH_2, SSE)$ is complete sufficient for $\underline{D}_1, \tau^2, \sigma^2$.

From (4.15),

$$
\begin{aligned}
SSH_1 &= tr\,[\,\hat{\underline{\beta}}_1^{\#T}(\underline{I} - \underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A})\hat{\underline{\beta}}_1^{\#}\underline{C}_{11}\,] \\
&= tr\,[\,\underline{C}_{11}^{1/2}\hat{\underline{\beta}}_1^{\#T}(\underline{I} - \underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A})\hat{\underline{\beta}}_1^{\#}\underline{C}_{11}^{1/2}\,],
\end{aligned}
$$

where $\underline{C}_{11}^{1/2}\hat{\underline{\beta}}_1^{\#T} \sim N_{p_1\otimes m}(\underline{C}_{11}^{1/2}\underline{D}_1\underline{A}^T,\ (\tau^2+\sigma^2)\underline{I}_{p_1} \otimes \underline{I}_m)$, and $(\underline{I} - \underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T)$ is idempotent with $rank = m - r$. Hence,

$$
\begin{aligned}
&\underline{C}_{11}^{1/2}\hat{\underline{\beta}}_1^{\#T}(\underline{I} - \underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A})\hat{\underline{\beta}}_1^{\#}\underline{C}_{11}^{1/2} \\
&\sim\ Wishart_{p_1}((\tau^2+\sigma^2)\underline{I}_{p_1}, m-r, \underline{0}). \qquad (4.16)
\end{aligned}
$$

By using (4.16), one gets

$$SSH_1 \sim (\tau^2 + \sigma^2)\chi^2_{(m-r)p_1}. \tag{4.17}$$

Due to the fact that $\hat{\underline{\beta}}_{k1}$'s are indepedent of $\tilde{\underline{\beta}}_{k2}$'s, (see Lemma 2 in Ghosh, Saleh, and Sen, 1987), and $SSH_2 \sim (\tau^2+\sigma^2)\chi^2_{mp_2}$, it is evident that $SSH_1 + SSH_2$ $\sim (\tau^2 + \sigma^2)\chi^2_{mp-rp_1}$. The $UMVUE$ of $\sigma^{-2}B = (\tau^2 + \sigma^2)^{-1}$ is then equivalent to $(mp - rp_1 - 2)/(SSH_1 + SSH_2)$ for $(mp - rp_1) \geq 3$ and $p = p_1 + p_2$. The best scale invariant estimate of σ^2 is $SSE/(N - mp + 2)$, where $N = \sum_{k=1}^{m} n_k$, and the MLE of $\underline{D}_1$ is $\hat{\underline{\beta}}_1^{\#T}\underline{A}(\underline{A}^T\underline{A})^{-1}$. Substituting all these estimators for the unknown parameters in (4.6), gives

$$\begin{aligned} & \underline{e}_{kEB}(\underline{Y}_k) \\ = \; & \underline{X}_1(\hat{\underline{D}}_1\underline{a}_k + (1 - \frac{mp - rp_1 - 2}{F})(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)) \\ + \; & (1 - \frac{mp - rp_1 - 2}{F})\underline{X}_2\tilde{\underline{\beta}}_{k2}, \; k = 1, \cdots, m, \end{aligned} \tag{4.18}$$

where $F = (N - mp + 2)(SSH_1 + SSH_2)/SSE$, a constant multiple of the F ratio for testing $H_0 : \underline{\beta}_{k1} = \underline{D}_1\underline{a}_k$ and $\underline{\beta}_{k2} = \underline{0}$ for all k.

In practice, it is much more appropriate to use the positive part EB estimator, $\underline{e}^{+*}_{EB} = (\underline{e}^{+T}_{1EB}, \cdots, \underline{e}^{+T}_{mEB})^T$, because B is estimated by the quantity $(mp - rp_1 - 2)/F$ which can be greater than 1.

$$\begin{aligned} & \underline{e}^{+}_{kEB}(\underline{Y}_k) \\ = \; & \underline{X}_1(\hat{\underline{D}}_1\underline{a}_k + (1 - \frac{mp - rp_1 - 2}{F})^{+}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)) \\ + \; & (1 - \frac{mp - rp_1 - 2}{F})^{+}\underline{X}_2\tilde{\underline{\beta}}_{k2}, \; k = 1, \cdots, m \end{aligned} \tag{4.19}$$

where $a^+ = \max(a, 0)$.

Note from (4.19) that for very large F values signifying substantial departure from $H_0 : \underline{\beta}_{k1} = \underline{D}_1\underline{a}_k$ and $\underline{\beta}_{k2} = \underline{0}$ for all k, $\underline{e}^{+}_{kEB}$ is very close to $\underline{X}\tilde{\underline{\beta}}_k$, whereas for very small F values signifying enough support for H_0, $\underline{e}^{+}_{kEB}$ is very close to $\underline{X}_1\hat{\underline{D}}_1\underline{a}_k$

for all k. When there is no clear-cut decision for or against H_0, $\underline{e}^{+}_{kEB}$ is in some sense a weighted average of $\underline{X}_1\hat{\underline{D}}_1\underline{a}_k$ and $\underline{X}\tilde{\underline{\beta}}_k$, with the weights being adaptively determined by the data.

Next, the estimators $\underline{e}^{*}_{EB} = (\underline{e}^{T}_{1EB}, \cdots, \underline{e}^{T}_{mEB})^T$ and $\underline{e}^{+*}_{EB}$ will be evaluated according to the criteria of frequentist and Bayes risks. First the frequentist criterion will be utilized.

The general class of estimators of $\underline{X\beta}_k$ is

$$\begin{aligned} \underline{e}_{k\phi}(\underline{Y}_k) &= \underline{X}_1(\hat{\underline{D}}_1\underline{a}_k + (1 - \frac{\phi(F)}{F})(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)) \\ &+ (1 - \frac{\phi(F)}{F})\underline{X}_2\tilde{\underline{\beta}}_{k2},\ k = 1, \cdots, m. \end{aligned} \tag{4.20}$$

The following theorem provides the risk expression for $\underline{e}^{*}_{\phi} = (\underline{e}^{T}_{1\phi}, \cdots, \underline{e}^{T}_{m\phi})^T$ under the loss (4.3) with $\underline{Q} = \underline{V}^{-1}$. The risk is denoted by $R(\underline{X\beta}_1, \cdots, \underline{X\beta}_m\ ; \underline{e}_{1\phi}, \cdots, \underline{e}_{m\phi})$.

Theorem 4.2.1. If ϕ is a differentiable function, then, under the loss (4.3) with $\underline{Q} = \underline{V}^{-1}$,

$$\begin{aligned} & R(\underline{X\beta}_k, \cdots, \underline{X\beta}_m; \underline{e}_{1\phi}, \cdots, \underline{e}_{m\phi}) \\ = \ & R(\underline{X\beta}_k, \cdots, \underline{X\beta}_m; \underline{X}\tilde{\underline{\beta}}_1, \cdots, \underline{X}\tilde{\underline{\beta}}_m) \\ - \ & 2\sigma^2 E_{\underline{\beta}_1, \cdots, \underline{\beta}_m, \sigma^2}\{\frac{\phi(F)}{F}(mp - rp_1) + 2(\frac{\phi'(F)}{F} - \frac{\phi(F)}{F^2})F\} \\ + \ & E_{\underline{\beta}_1, \cdots, \underline{\beta}_m, \sigma^2}\{\frac{\phi^2(F)}{F^2}(SSH_1 + SSH_2)\}. \end{aligned} \tag{4.21}$$

Proof of Theorem 4.2.1. By using (4.11), it can be shown that

$$\begin{aligned} & \underline{e}_{k\phi} - \underline{X\beta}_k \\ = \ & \underline{X}_1[\hat{\underline{\beta}}_{k1} - E(\hat{\underline{\beta}}_{k1}) - \frac{\phi(F)}{F}(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)\,] \\ + \ & (\underline{X}_2 - \underline{X}_1\underline{C}_{11}^{-1}\underline{C}_{12})[\,(1 - \frac{\phi(F)}{F})\tilde{\underline{\beta}}_{k2} - \underline{\beta}_{k2}\,]. \end{aligned} \tag{4.22}$$

Then,

$$
\begin{aligned}
&\sum_{k=1}^{m} E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}(\underline{e}_{k\phi} - \underline{X}\underline{\beta}_k)^T \underline{V}^{-1}(\underline{e}_{k\phi} - \underline{X}\underline{\beta}_k) \\
= &\sum_{k=1}^{m} E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{[\,\hat{\underline{\beta}}_{k1} - E(\hat{\underline{\beta}}_{k1}) - \frac{\phi(F)}{F}(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)\,]^T \\
&\underline{C}_{11}[\,\hat{\underline{\beta}}_{k1} - E(\hat{\underline{\beta}}_{k1}) - \frac{\phi(F)}{F}(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)\,] \\
+ &(\tilde{\underline{\beta}}_{k2} - \underline{\beta}_{k2} - \frac{\phi(F)}{F}\tilde{\underline{\beta}}_{k2})^T \underline{C}_{22.1}(\tilde{\underline{\beta}}_{k2} - \underline{\beta}_{k2} - \frac{\phi(F)}{F}\tilde{\underline{\beta}}_{k2})\} \\
= &\; mp\sigma^2 - 2E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{\frac{\phi(F)}{F}\sum_{k=1}^{m}[\,(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T \underline{C}_{11}(\hat{\underline{\beta}}_{k1} - E(\hat{\underline{\beta}}_{k1})) \\
+ &\; \tilde{\underline{\beta}}_{k2}^T \underline{C}_{22.1}(\tilde{\underline{\beta}}_{k2} - \underline{\beta}_{k2})\,]\} \\
+ &\; E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{\frac{\phi^2(F)}{F^2}\sum_{k=1}^{m}[\,(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T \underline{C}_{11}(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k) \\
+ &\; \tilde{\underline{\beta}}_{k2}^T \underline{C}_{22.1}\tilde{\underline{\beta}}_{k2}\,]\}.
\end{aligned}
\tag{4.23}
$$

A generalization of Stein's identity (Berger and Haff,1983) gives

$$
\begin{aligned}
&E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{\frac{\phi(F)}{F}\sum_{k=1}^{m}[\,(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T \underline{C}_{11}(\hat{\underline{\beta}}_{k1} - E(\hat{\underline{\beta}}_{k1})) \\
+ &\; \tilde{\underline{\beta}}_{k2}^T \underline{C}_{22.1}(\tilde{\underline{\beta}}_{k2} - \underline{\beta}_{k2})\,]\} \\
= &\sum_{k=1}^{m} E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{tr\,\{[\,\frac{\partial}{\partial\hat{\underline{\beta}}_{k1}}(\frac{\phi(F)}{F}(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T)\,]\underline{C}_{11}V(\hat{\underline{\beta}}_{k1})\} \\
+ &\; tr\,\{[\,\frac{\partial}{\partial\tilde{\underline{\beta}}_{k2}}(\frac{\phi(F)}{F}\tilde{\underline{\beta}}_{k2}^T)\,]\underline{C}_{22.1}V(\tilde{\underline{\beta}}_{k2})\}\} \\
= &\; \sigma^2 E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{2(\frac{\phi'(F)}{F} - \frac{\phi(F)}{F^2})\frac{SSH_1 + SSH_2}{SSE/(N-mp-2)} \\
+ &\; \frac{\phi(F)}{F}((m-r)p_1 + mp_2)\}.
\end{aligned}
\tag{4.24}
$$

Recall that $SSH_1 = \sum_{k=1}^{m}(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T\,\underline{C}_{11}(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)$, $SSH_2 = \sum_{k=1}^{m}\tilde{\underline{\beta}}_{k2}^T\,\underline{C}_{22.1}\tilde{\underline{\beta}}_{k2}$, and $F = ((SSH_1 + SSH_2)/SSE)(N - mp + 2)$. Since $R(\underline{X}\underline{\beta}_k, \cdots, \underline{X}\underline{\beta}_m; \underline{X}\tilde{\underline{\beta}}_1, \cdots, \underline{X}\tilde{\underline{\beta}}_m) = mp\sigma^2$, (4.21) can be obtained by combining (4.23) and (4.24). The proof of the theorem is complete.

The following corollary to Theorem 4.2.1 provides sufficient conditions under which $\underline{e}_\phi^*$ dominates $\underline{e}_{LSE}^* = ((\underline{X}\tilde{\underline{\beta}}_1)^T, \cdots, (\underline{X}\tilde{\underline{\beta}}_m)^T)^T$.

Corollary 4.2.1. Suppose $\underline{Q} = \underline{V}^{-1}$ and $mp - rp_1 \geq 3$; $\phi(F)$ is non-decreasing in F, and $0 < \phi(F) < 2(mp - rp_1 - 2)$. Then,

$$R(\underline{X\beta}_{\cdot 1}, \cdots, \underline{X\beta}_{\cdot m}; \underline{\varepsilon}_{1\phi}, \cdots, \underline{\varepsilon}_{m\phi}) < R(\underline{X\beta}_{\cdot 1}, \cdots, \underline{X\beta}_{\cdot m}; \underline{X\tilde{\beta}}_{\cdot 1}, \cdots, \underline{X\tilde{\beta}}_{\cdot m}).$$

Proof of Corollary 4.2.1. Corollary 4.2.1 can be obtained immediately by following the proof of Corollary 3.2.1.

Note that under the loss $\sigma^{-2}L(\underline{X\beta}_{\cdot 1}, \cdots, \underline{X\beta}_{\cdot m}; \underline{a}_1, \cdots, \underline{a}_m)$, $\underline{\varepsilon}^*_{LSE}$ is constant risk minimax estimator of $(\underline{X\beta})^*$. Therefore, any estimator dominating $\underline{\varepsilon}^*_{LSE}$ is also a minimax estimator of $(\underline{X\beta})^*$. Thus, Corollary 4.2.1 provides useful minimax estimators of $(\underline{X\beta})^*$. Also, it shows that $\underline{\varepsilon}^*_{EB}$ dominates $\underline{\varepsilon}^*_{LSE}$, and is a minimax estimator if $\underline{Q} = \underline{V}^{-1}$ for $\underline{\varepsilon}^*_{EB}$ being a special case of $\underline{\varepsilon}^*_{\phi}$ with $\phi(F) = mp - rp_1 - 2$. It is clear from (4.21) that when $\phi(F) = c$, a positive constant, then $R(\underline{X\beta}_{\cdot 1}, \cdots, \underline{X\beta}_{\cdot m}; \underline{\varepsilon}_{1\phi}, \cdots, \underline{\varepsilon}_{m\phi})$ is minimized when $c = mp - rp_1 - 2$. Thus, for $\underline{Q} = \underline{V}^{-1}$, $\underline{\varepsilon}^*_{EB}$ is optimal within the class of estimation of the form

$$\underline{X}_1(\hat{\underline{D}}_1\underline{a}_k + (1 - \frac{c}{F})(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)) + (1 - \frac{c}{F})\underline{X}_2\tilde{\underline{\beta}}_{k2},\ k = 1, \cdots, m.$$

It will be shown later that the Bayesian opatimality of $\underline{\varepsilon}^*_{EB}$ holds within the same class of estimators irrespective of any p.d. matrix $\underline{Q}$.

It is not necessarily true that $\underline{\varepsilon}^*_{EB}$ dominate $\underline{\varepsilon}^*_{RMLE} = ((\underline{X}_1\hat{\underline{D}}_1\underline{a}_1)^T, \cdots, (\underline{X}_1\hat{\underline{D}}_1\underline{a}_m)^T)^T$. By expressing

$$\begin{aligned} & \underline{X}_1\hat{\underline{D}}_1\underline{a}_k - \underline{X\beta}_{\cdot k} \\ = \ & \underline{X}_1(\hat{\underline{D}}_1\underline{a}_k - E(\hat{\underline{D}}_1\underline{a}_k)) + \underline{X}_1(E(\hat{\underline{D}}_1\underline{a}_k) - E(\hat{\underline{\beta}}_{k1})) \\ + \ & (\underline{X}_1\underline{C}_{11}^{-1}\underline{C}_{12} - \underline{X}_2)\underline{\beta}_{k2}, \end{aligned}$$

it is easy to check that in estimating $(\underline{X\beta})^*$ $\underline{\varepsilon}^*_{RMLE}$ has the risk

$$\begin{aligned} & R(\underline{X\beta}_{\cdot 1}, \cdots, \underline{X\beta}_{\cdot m}; \underline{X}_1\hat{\underline{D}}_1\underline{a}_1, \cdots, \underline{X}_1\hat{\underline{D}}_1\underline{a}_m) \\ = \ & \sum_{k=1}^{m} tr\,[\,\underline{C}_{11}V(\hat{\underline{D}}_1\underline{a}_k)\,] + \sum_{k=1}^{m} E(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T\underline{C}_{11}E(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k) \\ + \ & \sum_{k=1}^{m} \underline{\beta}_{k2}^T\underline{C}_{22.1}\underline{\beta}_{k2}, \end{aligned} \tag{4.25}$$

where $\sum_{k=1}^{m} tr\,[\,\underline{C}_{11}V(\underline{\hat{D}}_1\underline{a}_k)\,] = \sigma^2 p_1 \sum_{k=1}^{m} \underline{a}_k^T(\underline{A}^T\underline{A})^{-1}\underline{a}_k = \sigma^2 p_1 tr(\underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T) = rp_1\sigma^2$.

Also, from (4.21) with $\phi(F) = mp - rp_1 - 2$, it can be obtained that

$$R(\underline{X\beta}_1,\cdots,\underline{X\beta}_m;\underline{e}_{1EB},\cdots,\underline{e}_{mEB}) = mp\sigma^2$$
$$- \ 2\sigma^2(mp - rp_1 - 2)^2E(\frac{1}{F}) + (mp - rp_1 - 2)^2E(\frac{SSH_1 + SSH_2}{F^2}). \quad (4.26)$$

For fixed $\underline{\beta}_1,\cdots,\underline{\beta}_m$, and σ^2, the variables SSH_1, SSH_2, and SSE are independently distributed with $(SSH_1 + SSH_2) \sim \sigma^2\ \chi^2_{mp-rp_1}(\lambda)$, $\lambda = (1/(2\sigma^2))(\sum_{k=1}^{m}$ $[\,E(\underline{\hat{\beta}}_{k1} - \underline{\hat{D}}_1\underline{a}_k)^T\ \underline{C}_{11}\ E(\underline{\hat{\beta}}_{k1} - \underline{\hat{D}}_1\underline{a}_k) + \sum_{k=1}^{m}\ \underline{\beta}_{k2}^T\underline{C}_{22.1}\ \underline{\beta}_{k2}\,]$, and $SSE \sim \sigma^2\chi^2_{N-mp}$. It follows from (4.26) that

$$R(\underline{X\beta}_1,\cdots,\underline{X\beta}_m;\underline{e}_{1EB},\cdots,\underline{e}_{mEB})$$
$$= \ mp\sigma^2 - \sigma^2(mp - rp_1 - 2)^2(\frac{N - mp}{N - mp + 2})E(mp - rp_1 - 2 + 2K)^{-1}, \quad (4.27)$$

where $K \sim Poisson(\lambda)$. Thus from (4.25) and (4.27), as $\lambda \to 0$, $R(\underline{X\beta}_1,\cdots,\ \underline{X\beta}_m;$ $\underline{X}_1\underline{\hat{D}}_1\underline{a}_1,\cdots,\underline{X}_1\underline{\hat{D}}_1\underline{a}_m) \to rp_1\sigma^2$, in which $R(\underline{X\beta}_1,\cdots,\underline{X\beta}_m;\ \underline{e}_{1EB},\cdots,\underline{e}_{mEB}) \to \sigma^2$ $[\,mp - (mp - rp_1 - 2)(N - mp)/(N - mp + 2)\,] > rp_1\sigma^2$. On the other hand, as $\lambda \to \infty$, $R(\underline{X\beta}_1,\cdots,\ \underline{X\beta}_m;\ \underline{X}_1\underline{\hat{D}}_1\underline{a}_1,\cdots,\ \underline{X}_1\underline{\hat{D}}_1\underline{a}_m) \to \infty$, but $R(\underline{X\beta}_1,\ \cdots,\underline{X\beta}_m;$ $\underline{e}_{1EB},\ \cdots,\underline{e}_{mEB}) \to mp\sigma^2$. Hence neither $\underline{e}^*_{RMLE}$ nor $\underline{e}^*_{EB}$ domintates each other. The above phenomenon is expected. Small λ indicates that $\underline{\beta}_{k1}$ is close to $\underline{D}_1\underline{a}_k$, and $\underline{\beta}_{k2}$ is close to $\underline{0}$ for all k, in which case clearly $\underline{e}^*_{RMLE}$ is the desired estimator. On the other hand, large λ indicates substantial rejection of $H_0 : \underline{\beta}_{k1} = \underline{D}_1\underline{a}_k$ and $\underline{\beta}_{k2} = \underline{0}$ for all k, in which case the model (4.1) is expected to perform well. The EB estimator turns out to be fairly robust. Note that $\underline{e}^*_{RMLE}$ is not a minimax estimator of $(\underline{X\beta})^*$. Since $R(\underline{X\beta}_1,\cdots,\ \underline{X\beta}_m;\ \underline{X}_1\underline{\hat{D}}_1\underline{a}_1,\cdots,\ \underline{X}_1\underline{\hat{D}}_1\underline{a}_m) \to \infty$ as $\lambda \to \infty$, from the robustness criterion, $\underline{e}^*_{EB}$ is evidently the superior over $\underline{e}^*_{LSE}$ or $\underline{e}^*_{RMLE}$ when $\underline{Q} = \underline{V}^{-1}$.

However, the estimator $\underline{e}^*_{EB}$ is obtained by estimating B, the Bayes shrinking factor, by a random variable which can be bigger than 1 with positive probability. As mentioned earlier, this deficiency is remedied by $\underline{e}^{+*}_{EB}$. For the every estimator $\underline{e}^*_\phi$ defined in (4.20), there is a corresponding estimator $\underline{e}^{+*}_\phi$ given by

$$\begin{aligned}\underline{e}_{k\phi}(\underline{Y}_k) &= \underline{X}_1(\hat{\underline{D}}_1\underline{a}_k + (1-\frac{\phi(F)}{F})^+(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)) \\ &+ (1-\frac{\phi(F)}{F})^+\underline{X}_2\tilde{\underline{\beta}}_{k2},\ k=1,\cdots,m. \qquad (4.28)\end{aligned}$$

It can be expected that $\underline{e}^{+*}_\phi$ has smaller risk than $\underline{e}^*_\phi$.

Theorem 4.2.2. Under the loss (4.3) with $\underline{Q} = \underline{V}^{-1}$,

$$R(\underline{X\beta}_1,\cdots,\underline{X\beta}_m;\underline{e}^+_{1\phi},\cdots,\underline{e}^+_{m\phi}) < R(\underline{X\beta}_1,\cdots,\underline{X\beta}_m;\underline{e}_{1\phi},\cdots,\underline{e}_{m\phi}). \qquad (4.29)$$

Proof of Theorem 4.2.2. After some algebra, one gets

$$\begin{aligned}&R(\underline{X\beta}_1,\cdots,\underline{X\beta}_m;\underline{e}_{1\phi},\cdots,\underline{e}_{m\phi}) - R(\underline{X\beta}_1,\cdots,\underline{X\beta}_m;\underline{e}^+_{1\phi},\cdots,\underline{e}^+_{m\phi}) \\ =\ &\sum_{k=1}^m E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{[\,(1-\frac{\phi(F)}{F})^2 - ((1-\frac{\phi(F)}{F})^+)^2\,] \\ &(\underline{X}_1(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k) + \underline{X}_2\tilde{\underline{\beta}}_{k2})^T\underline{V}^{-1}(\underline{X}_1(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k) + \underline{X}_2\tilde{\underline{\beta}}_{k2}) \\ +\ &2[\,(1-\frac{\phi(F)}{F}) - (1-\frac{\phi(F)}{F})^+\,] \\ &(\underline{X}_1\hat{\underline{D}}_1\underline{a}_k - \underline{X\beta}_k)^T\underline{V}^{-1}(\underline{X}_1(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k) + \underline{X}_2\tilde{\underline{\beta}}_{k2})\} \\ \geq\ &\sum_{k=1}^m -2E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{(1-\frac{\phi(F)}{F})I_{[\frac{\phi(F)}{F}>1]} \\ &(\underline{X}_1\hat{\underline{D}}_1\underline{a}_k - \underline{X\beta}_k)^T\underline{V}^{-1}(\underline{X}_1\hat{\underline{D}}_1\underline{a}_k - \underline{X}\tilde{\underline{\beta}}_k)\}. \qquad (4.30)\end{aligned}$$

Using the identities

$$\begin{aligned}&\underline{X}_1\hat{\underline{D}}_1\underline{a}_k - \underline{X\beta}_k = \underline{X}_1(\hat{\underline{D}}_1\underline{a}_k - E(\hat{\underline{D}}_1\underline{a}_k)), \\ +\ &\underline{X}_1(E(\hat{\underline{D}}_1\underline{a}_k) - E(\hat{\underline{\beta}}_{k1})) + (\underline{X}_1\underline{C}_{11}^{-1}\underline{C}_{12} - \underline{X}_2)\underline{\beta}_{k2}, \qquad (4.31)\end{aligned}$$

and

$$\begin{aligned}&\underline{X}_1\hat{\underline{D}}_1\underline{a}_k - \underline{X}\tilde{\underline{\beta}}_k \\ =\ &\underline{X}_1(\hat{\underline{D}}_1\underline{a}_k - \hat{\underline{\beta}}_{k1}) + (\underline{X}_1\underline{C}_{11}^{-1}\underline{C}_{12} - \underline{X}_2)\tilde{\underline{\beta}}_{k2}, \qquad (4.32)\end{aligned}$$

and the mutual independence of $\hat{\underline{D}}_1$ with $(\hat{\underline{D}}_1\underline{a}_1 - \hat{\underline{\beta}}_{11}, \cdots, \hat{\underline{D}}_1\underline{a}_m - \hat{\underline{\beta}}_{m1})$, SSH_2, and SSE. Recall that $SSH_1 = \sum_{k=1}^{m}(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T \underline{C}_{11}(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)$ and F is a function of SSH_1, SSH_2 and SSE.

The right hand side of (4.30) can be expressed as

$$\begin{aligned} & \sum_{k=1}^{m} E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{(E(\hat{\underline{D}}_1\underline{a}_k - \hat{\underline{\beta}}_{k1}))^T \underline{C}_{11}(\hat{\underline{D}}_1\underline{a}_k - \hat{\underline{\beta}}_{k1}) \\ + \ & \underline{\beta}_{k2}^T \underline{C}_{22.1}\tilde{\underline{\beta}}_{k2}\}(1 - \frac{\phi(F)}{F})I_{[\frac{\phi(F)}{F}>1]}. \end{aligned} \tag{4.33}$$

Hence, it suffices to show that

$$\begin{aligned} & \sum_{k=1}^{m} E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{(E(\hat{\underline{D}}_1\underline{a}_k - \hat{\underline{\beta}}_{k1}))^T \underline{C}_{11}(\hat{\underline{D}}_1\underline{a}_k - \hat{\underline{\beta}}_{k1}) \\ + \ & \underline{\beta}_{k2}^T \underline{C}_{22.1}\tilde{\underline{\beta}}_{k2} | SSH_1 = h_1, \\ & \tilde{\underline{\beta}}_{k2}^T \underline{C}_{22.1}\tilde{\underline{\beta}}_{k2} = h_{2k}, \text{ for all } k = 1, \cdots, m, \\ & SSE = e, \exists \frac{\phi(F)}{F} > 1\} > 0, \end{aligned} \tag{4.34}$$

for all $h_1 > 0, h_{2k} > 0$ and $e > 0$. Since $(\hat{\underline{\beta}}_{11}^T \cdots \hat{\underline{\beta}}_{m1}^T)^T$, $(\tilde{\underline{\beta}}_{12}^T \cdots \tilde{\underline{\beta}}_{m2}^T)^T$, and SSE are mutually independent, (4.34) can be then written as

$$\begin{aligned} & \sum_{k=1}^{m} E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{\{(E(\hat{\underline{D}}_1\underline{a}_k - \hat{\underline{\beta}}_{k1}))^T \underline{C}_{11}(\hat{\underline{D}}_1\underline{a}_k - \hat{\underline{\beta}}_{k1}) | SSH_1 = h_1) \\ + \ & (\underline{\beta}_{k2}^T \underline{C}_{22.1}\tilde{\underline{\beta}}_{k2} | \tilde{\underline{\beta}}_{k2}^T \underline{C}_{22.1}\tilde{\underline{\beta}}_{k2} = h_{2k})\} > 0, \text{ for all } h_1, h_{2k} > 0. \end{aligned} \tag{4.35}$$

From the proof of Theorem 2.2.3, it can be shown that

$$\begin{aligned} & \sum_{k=1}^{m} E_{\underline{\beta}_k,\sigma^2}\{(\underline{\beta}_{k2}^T \underline{C}_{22.1}\tilde{\underline{\beta}}_{k2} | \tilde{\underline{\beta}}_{k2}^T \underline{C}_{22.1}\tilde{\underline{\beta}}_{k2} = h_{2k}\} > 0, \\ & \text{for every } h_{2k} > 0. \end{aligned} \tag{4.36}$$

Recall SSH_1 can be expressed as

$$tr\,[\,\underline{C}_{11}^{1/2}\hat{\underline{\beta}}_1^{\#T}(\underline{I} - \underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T)\hat{\underline{\beta}}_1^{\#}\underline{C}_{11}^{1/2}\,].$$

Thus, the former term of (4.35) can be shown as

$$\begin{aligned} & E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{tr\,[\,\underline{C}_{11}^{1/2}E(\hat{\underline{\beta}}_1^{\#T})(\underline{I} - \underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T)\hat{\underline{\beta}}_1^{\#}\underline{C}_{11}^{1/2}\,] \\ & | tr\,[\,\underline{C}_{11}^{1/2}\hat{\underline{\beta}}_1^{\#T}(\underline{I} - \underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T)\hat{\underline{\beta}}_1^{\#}\underline{C}_{11}^{1/2}\,] = h_1\}. \end{aligned} \tag{4.37}$$

Now, let

$$(\underline{\alpha}_1, \cdots, \underline{\alpha}_m) = \underline{C}_{11}^{1/2}(\underline{\hat{\beta}}_{11}, \cdots, \underline{\hat{\beta}}_{m1}). \tag{4.38}$$

The distribution of $(\underline{\alpha}_1, \cdots, \underline{\alpha}_m)$ is $N_{p_1 \otimes m}(\underline{C}_{11}^{1/2} E(\underline{\hat{\beta}}_{11}, \cdots, \underline{\hat{\beta}}_{m1}), \sigma^2 \underline{I}_{p_1} \otimes \underline{I}_m)$. There exists an orthogonal matrix $\underline{P}_{m \times m}$ such that

$$\underline{P}^T(\underline{I} - \underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T)\underline{P} = \begin{bmatrix} \underline{I}_{m-r} & \underline{0}_{(m-r)\times r} \\ \underline{0}_{r\times(m-r)} & \underline{0}_{r\times r} \end{bmatrix}. \tag{4.39}$$

From (4.38) and (4.39), the latter part of (4.37) can be expressed as

$$tr\{[\underline{\phi}_1, \cdots, \underline{\phi}_m] \begin{bmatrix} \underline{I}_{m-r} & \underline{0}_{(m-r)\times r} \\ \underline{0}_{r\times(m-r)} & \underline{0}_{r\times r} \end{bmatrix} \begin{bmatrix} \underline{\phi}_1^T \\ \vdots \\ \underline{\phi}_m^T \end{bmatrix}\}, \tag{4.40}$$

where $(\underline{\phi}_1, \cdots, \underline{\phi}_m) = (\underline{\alpha}_1, \cdots, \underline{\alpha}_m)\underline{P}$ and is distributed as $N_{p_1 \otimes m}(\underline{C}_{11}^{1/2} E(\underline{\hat{\beta}}_1^{\#T})\underline{P}, \sigma^2 \underline{I}_{p_1} \otimes \underline{I}_m)$. Now, (4.40) can be rewritten as

$$\begin{aligned} & tr\{\begin{bmatrix} \underline{\phi}_1^T \\ \vdots \\ \underline{\phi}_m^T \end{bmatrix} [\underline{\phi}_1, \cdots, \underline{\phi}_m] \begin{bmatrix} \underline{I}_{m-r} & \underline{0}_{(m-r)\times r} \\ \underline{0}_{r\times(m-r)} & \underline{0}_{r\times r} \end{bmatrix}\} \\ = \; & tr\{\begin{bmatrix} \underline{U}_{(m-r)} & \underline{0}_{(m-r)\times r} \\ \underline{0}_{r\times(m-r)} & \underline{0}_r \end{bmatrix}\} = \sum_{k=1}^{m-r} \underline{\phi}_k^T \underline{\phi}_k, \end{aligned} \tag{4.41}$$

where $\underline{U} = Diag(\underline{\phi}_1^T\underline{\phi}_1, \cdots, \underline{\phi}_{m-r}^T\underline{\phi}_{m-r})$.

Since $\underline{\phi}_k$'s are independent normal vectors, each has the mean equal to the kth column of $\underline{C}_{11}^{1/2} E(\underline{\hat{\beta}}_1^{\#T})\underline{P}$ and the variance-covariance matrix equal to $\underline{I}_{p_1}$, and from (4.41), (4.37) can be then expressed as

$$\begin{aligned} & \sum_{k=1}^{m-r} [\, (E(\underline{\phi}_k))^T \underline{\phi}_k | \sum_{j=1}^{m-r} \underline{\phi}_j^T \underline{\phi}_j \,] \\ = \; & \sum_{k=1}^{m-r} [\, (E(\underline{\phi}_k))^T \underline{\phi}_k | \underline{\phi}_k^T \underline{\phi}_k \,] \geq 0. \end{aligned} \tag{4.42}$$

The previous statement again is based on the argument in Theorem 6.2 of Lehmann (1983). Theorem 4.2.2 is hence proved as a consequence from (4.36) and (4.42).

Finally, based on the hypothesis $H_0 : \underline{\beta}_{k1} = \underline{D}_1 \underline{a}_k$ and $\underline{\beta}_{k2} = \underline{0}$ for all k, a general PTE is given by

$$\underline{\delta}_{kPTE} = \underline{X}_1(\underline{\hat{D}}_1 \underline{a}_k + g(F)(\underline{\tilde{\beta}}_{k1} - \underline{\hat{D}}_1 \underline{a}_k)) + \underline{X}_2 g(F) \underline{\tilde{\beta}}_{k2}, \tag{4.43}$$

where $g(F) = I_{[F>d]}$ with d, a postive constant depending upon the chosen level of significance, and I the usual indicator function. The corresponding modified EB estimator,

$$\underline{\hat{e}}^{\phi_0}_{kMEB}(\underline{Y}_k) = \underline{X}_1(\underline{\hat{D}}_1\underline{a}_k + (1 - \frac{\phi_0(F)}{F})g(F)(\underline{\tilde{\beta}}_{k1} - \underline{\hat{D}}_1\underline{a}_k)) + \underline{X}_2(1 - \frac{\phi_0(F)}{F})g(F)\underline{\tilde{\beta}}_{k2},\ k = 1, \cdots, m, \tag{4.44}$$

dominates $\underline{\delta}^*_{PTE} = (\underline{\delta}^T_{1PTE}, \cdots, \underline{\delta}^T_{mPTE})^T$ under certain conditions.

Theorem 4.2.3. Consider the loss (4.3) with $\underline{Q} = \underline{V}^{-1}$. If $0 < \phi_0(F) < 2(mp - rp_1 - 2)$, then

$$\begin{aligned} & R(\underline{X\beta}_1, \cdots, \underline{X\beta}_m; \underline{\delta}^{\phi_0}_{1MEB}, \cdots, \underline{\delta}^{\phi_0}_{mMEB}) \\ < \ & R(\underline{X\beta}_1, \cdots, \underline{X\beta}_m; \underline{\delta}_{1PTE}, \cdots, \underline{\delta}_{mPTE}). \end{aligned} \tag{4.45}$$

Proof of Theorem 4.2.3. By using (4.21), and following the procedure in Proof of Theorem 2.2.4, (4.45) can be obtained immediately.

Note that $\underline{\delta}^{\phi_0}_{kMEB}$ estimates $\underline{X\beta}_k$ by $\underline{X}_1\underline{\hat{D}}_1\underline{a}_k$ if $g(F) = 0$, i.e., the null hypothesis is accepted at a desired level of significance, and it estimates $\underline{X\beta}_k$ by an EB type estimator as described earlier in this section when the other is true. Also, the modified EB estimator $\underline{\delta}^{\phi_0*}_{MEB}$=$(\underline{\delta}^{\phi_0 T}_{1MEB}, \cdots, \underline{\delta}^{\phi_0 T}_{mMEB})^T$ can be further improved by a positive part modified estimator $(\underline{\delta}^{\phi_0}_{MEB})^{+*} = (\underline{\delta}^{\phi_0^+ T}_{1MEB}, \cdots, \underline{\delta}^{\phi_0^+ T}_{mMEB})^T$ as

$$\begin{aligned} (\underline{\beta}^{\phi_0}_{kMEB}(\underline{Y}_k))^+ = \ & \underline{X}_1(\underline{\hat{D}}_1\underline{a}_k + (1 - \frac{\phi_0(F)}{F})^+ g(F)(\underline{\tilde{\beta}}_{k1} - \underline{\hat{D}}_1\underline{a}_k)) \\ + \ & \underline{X}_2(1 - \frac{\phi_0(F)}{F})^+ g(F)\underline{\tilde{\beta}}_{k2},\ k = 1, \cdots, m. \end{aligned} \tag{4.46}$$

Now, $\underline{e}^*_{EB}$ and other associated estimators of $(\underline{X\beta})^*$ are compared by the Bayes risk. In view of Theorem 4.2.1, the Bayes risk of $\underline{e}^*_{EB}$ is smaller than that of $\underline{e}^*_{LSE}$ when $\underline{Q} = \underline{V}^{-1}$ without regarding of any prior, while $\underline{e}^+_{kEB}$ always dominates $\underline{e}_{kEB}$ irrespective of any prior. However, the Bayes risk optimatity of $\underline{e}^*_{EB}$ over $\underline{e}^*_{LSE}$ is

not limited to $\underline{Q} = \underline{V}^{-1}$. The following theorem will show this superiority of $\underline{\varepsilon}^*_{EB}$ over $\underline{\varepsilon}^*_{LSE}$ under the prior presented in the beginning of this section.

Theorem 4.2.4. Consider the model $\underline{Y}_k|\underline{\beta}_k \sim N(\underline{X}\underline{\beta}_k, \sigma^2\underline{V})$, for $k = 1, \cdots, m$, where $\underline{V}$ is p.d., and $\underline{\beta}_k \sim N(\begin{pmatrix} \underline{D}_1\underline{a}_k \\ \underline{0} \end{pmatrix}, \tau^2(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1})$ for all k. Let ξ denote this given prior, and $r(\xi; \underline{\varepsilon}_1, \cdots, \underline{\varepsilon}_m)$ the Bayes risk of an estimator $(\underline{\varepsilon}_1, \cdots, \underline{\varepsilon}_m)$ of $(\underline{X}\underline{\beta}_1, \cdots, \underline{X}\underline{\beta}_m)$ under the prior ξ when the loss is given in (4.3). Then, it is clear that

$$r(\xi; \underline{\varepsilon}_{1EB}, \cdots, \underline{\varepsilon}_{mEB}) < r(\xi; \underline{X}\tilde{\underline{\beta}}_1, \cdots, \underline{X}\tilde{\underline{\beta}}_m). \tag{4.47}$$

Proof of Theorem 4.2.4. Let

$$\underline{\varepsilon}_{kc} = \underline{X}_1(\tilde{\underline{\beta}}_{k1} - \frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)) + (1 - \frac{c}{F})\underline{X}_2\tilde{\underline{\beta}}_{k2},\ k = 1, \cdots, m. \tag{4.48}$$

According to the loss (4.3),

$$\begin{aligned}
& r(\xi; \underline{\varepsilon}_{1c}, \cdots, \underline{\varepsilon}_{mc}) \\
= & \sum_{k=1}^{m} tr[\{E(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1} - \frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)) \\
& (\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1} - \frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k))^T\}\underline{X}_1^T\underline{Q}\underline{X}_1] \\
+ & \sum_{k=1}^{m} tr[\{E(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1} - \frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)) \\
& (\tilde{\underline{\beta}}_{k2} - \underline{\beta}_{k2} - \frac{c}{F}\tilde{\underline{\beta}}_{k2})^T\}\underline{X}_2^T\underline{Q}\underline{X}_1] \\
+ & \sum_{k=1}^{m} tr[\{E(\tilde{\underline{\beta}}_{k2} - \underline{\beta}_{k2} - \frac{c}{F}\tilde{\underline{\beta}}_{k2}) \\
& (\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1} - \frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k))^T\}\underline{X}_1^T\underline{Q}\underline{X}_2] \\
+ & \sum_{k=1}^{m} tr[\ E(\tilde{\underline{\beta}}_{k2} - \underline{\beta}_{k2} - \frac{c}{F}\tilde{\underline{\beta}}_{k2})(\tilde{\underline{\beta}}_{k2} - \underline{\beta}_{k2} - \frac{c}{F}\tilde{\underline{\beta}}_{k2})^T\}\underline{X}_2^T\underline{Q}\underline{X}_2],
\end{aligned} \tag{4.49}$$

in which

$$\begin{aligned}
& E(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1} - \frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k))(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1} - \frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k))^T \\
= & \ \sigma^2\underline{C}_{11.2}^{-1} + E(\frac{c^2}{F^2}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T) \\
- & \ E(\frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1})(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T) - E(\frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1})^T),
\end{aligned} \tag{4.50}$$

and

$$
\begin{aligned}
& E\Big(\frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1})(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T\Big) \\
= \; & \sigma^2 \frac{c(N-mp)}{N-mp+2} E\Big(\frac{1}{SSH_1 + SSH_2}(\tilde{\underline{\beta}}_{k1} - E(\underline{\beta}_{k1}|\underline{Y}_k)) \\
& (\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T\Big). \qquad (4.51)
\end{aligned}
$$

By using (4.5), (4.51) can be expressed as

$$
\begin{aligned}
& \sigma^2 B \frac{c(N-mp)}{N-mp+2} E\Big(\frac{1}{SSH_1 + SSH_2}\{(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T \\
+ \; & (\tilde{\underline{D}}_1\underline{a}_k - \underline{D}_1\underline{a}_k)(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T\}\Big). \qquad (4.52)
\end{aligned}
$$

By using (4.11), the independence between $\hat{\underline{\beta}}_{k1}$'s and $\tilde{\underline{\beta}}_{k2}$'s, the independence between $\hat{\underline{D}}_1\underline{a}_k$ and SSH_1, and that between $\hat{\underline{D}}_1\underline{a}_k$ and $\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k$, and also $E\hat{\underline{D}}_1 = \underline{D}_1$, from Basu's theorem, and using Lemma 1 in Ghosh, Saleh, and Sen (1987), Equation (4.52) equals to

$$
\begin{aligned}
& \sigma^2 B \frac{c(N-mp)}{N-mp+2}\Big\{\frac{V(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)}{E(SSH_1 + SSH_2)} + \frac{\underline{C}_{11}^{-1}\underline{C}_{12}V(\tilde{\underline{\beta}}_{k2})\underline{C}_{21}\underline{C}_{11}^{-1}}{E(SSH_1 + SSH_2)}\Big\} \\
= \; & \sigma^2 B \frac{c(N-mp)}{(N-mp+2)(mp-rp_1)}(\underline{C}_{11.2}^{-1} - (\underline{a}_k^T(\underline{A}^T\underline{A})^{-1}\underline{a}_k)\underline{C}_{11}^{-1}), \qquad (4.53)
\end{aligned}
$$

According to the same argument in (4.53),

$$
\begin{aligned}
& E\Big(\frac{c^2}{F^2}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T\Big) \\
= \; & \sigma^2 B \frac{c^2(N-mp)}{N-mp+2}\frac{(\underline{C}_{11.2}^{-1} - (\underline{a}_k^T(\underline{A}^T\underline{A})^{-1}\underline{a}_k)\underline{C}_{11}^{-1})}{(mp-rp_1)(mp-rp_1-2)}. \qquad (4.54)
\end{aligned}
$$

By combining (4.50), (4.53) and (4.54), the first term of (4.49) becomes

$$
\begin{aligned}
& m\{\sigma^2 tr\,[\,\underline{C}_{11.2}^{-1}(\underline{X}_1^T\underline{Q}\underline{X}_1)\,]\} + \Big(\frac{c^2}{mp-rp_1-2} - 2c\Big)\sigma^2 B\frac{N-mp}{(N-mp+2)(mp-rp_1)} \\
& tr\,[\,((m-r)\underline{C}_{11}^{-1} + m\underline{C}_{11}^{-1}\underline{C}_{12}\underline{C}_{22.1}^{-1}\underline{C}_{21}\underline{C}_{11}^{-1})\underline{X}_1^T\underline{Q}\underline{X}_1\,]. \qquad (4.55)
\end{aligned}
$$

Now, applying Basu's theorem and Lemma 1 in Ghosh, Salen, and Sen (1987),

$$
\begin{aligned}
& E(\tilde{\underline{\beta}}_{k2} - \underline{\beta}_{k2} - \frac{c}{F}\tilde{\underline{\beta}}_{k2})(\tilde{\underline{\beta}}_{k2} - \underline{\beta}_{k2} - \frac{c}{F}\tilde{\underline{\beta}}_{k2})^T \\
= \; & -\sigma^2\underline{C}_{22.1}^{-1} - E\Big(\frac{c}{F}\tilde{\underline{\beta}}_{k2}(\tilde{\underline{\beta}}_{k2} - E(\underline{\beta}_{k2}|\underline{Y}_k))^T\Big) + E\Big(\frac{c^2}{F^2}\tilde{\underline{\beta}}_{k2}\tilde{\underline{\beta}}_{k2}^T\Big) \\
= \; & \sigma^2\underline{C}_{22.1}^{-1} + \Big(\frac{c^2}{mp-rp_1-2} - 2c\Big)\sigma^2 B\frac{N-mp}{(N-mp+2)(mp-rp_1)}\underline{C}_{22.1}^{-1}. \qquad (4.56)
\end{aligned}
$$

From (4.56), the last term of (4.49) is identical to

$$m\{\sigma^2 tr\,[\,\underline{C}_{22.1}^{-1}(\underline{X}_2^T\underline{Q}\underline{X}_2)\,]\} + (\frac{c^2}{mp-rp_1-2} - 2c)\sigma^2 B\frac{N-mp}{(N-mp+2)(mp-rp_1)}$$
$$tr\,[\,m\underline{C}_{22.1}^{-1}\underline{X}_2^T\underline{Q}\underline{X}_2\,]. \tag{4.57}$$

By using the same argument, the cross-product term becomes

$$\begin{aligned}
& E(\tilde{\underline{\beta}}_{k1} - \underline{\beta}_{k1} - \frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k))(\tilde{\underline{\beta}}_{k2} - \underline{\beta}_{k2} - \frac{c}{F}\tilde{\underline{\beta}}_{k2})^T \\
= \;& \sigma^2(\underline{C}_{11.2}^{-1}\underline{C}_{12}\underline{C}_{22}^{-1}) - E(\frac{c}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)(\tilde{\underline{\beta}}_{k2} - E(\underline{\beta}_{k2}|\underline{Y}_k))^T) \\
- \;& E(\frac{c}{F}(\tilde{\underline{\beta}}_{k1} - E(\underline{\beta}_{k1}|\underline{Y}_k))\tilde{\underline{\beta}}_{k2}^T) + E(\frac{c^2}{F^2}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)\tilde{\underline{\beta}}_{k2}^T) \\
= \;& -\sigma^2(\underline{C}_{11.2}^{-1}\underline{C}_{12}\underline{C}_{22}^{-1}) \\
- \;& (\frac{c^2}{mp-rp_1-2} - 2c)\sigma^2 B\frac{N-mp}{(N-mp+2)(mp-rp_1)}\underline{C}_{11}^{-1}\underline{C}_{12}\underline{C}_{22.1}^{-1}.
\end{aligned} \tag{4.58}$$

From (4.58), the cross-product term in (4.49) can be expressed as

$$\begin{aligned}
& \{-m\sigma^2 tr\,[\,\underline{C}_{11.2}^{-1}\underline{C}_{12}\underline{C}_{22}^{-1}\underline{X}_2^T\underline{Q}\underline{X}_1\,]\} \\
- \;& (\frac{c^2}{mp-rp_1-2} - 2c)\sigma^2 B\frac{N-mp}{(N-mp+2)(mp-rp_1)}tr\,[\,m\underline{C}_{11}^{-1}\underline{C}_{12}\underline{C}_{22.1}^{-1}\underline{X}_2^T\underline{Q}\underline{X}_1\,] \\
- \;& m\sigma^2 tr\,[\,\underline{C}_{22.1}^{-1}\underline{C}_{21}\underline{C}_{11}^{-1}\underline{X}_1^T\underline{Q}\underline{X}_2\,] - (\frac{c^2}{mp-rp_1-2} - 2c)\sigma^2 B\frac{N-mp}{(N-mp+2)(mp-rp_1)} \\
& tr\,[\,m\underline{C}_{22}^{-1}\underline{C}_{21}\underline{C}_{11.2}^{-1}\underline{X}_1^T\underline{Q}\underline{X}_2\,]\}.
\end{aligned} \tag{4.59}$$

Finally, by combining (4.55), (4.57) and (4.59), it can be obtained that

$$\begin{aligned}
& r(\xi;\underline{e}_{1c},\cdots,\underline{e}_{mc}) \\
= \;& m\sigma^2 tr\,[\,(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\underline{X}^T\underline{Q}\underline{X}\,] \\
+ \;& (\frac{c^2}{mp-rp_1-2} - 2c)\sigma^2 B\frac{N-mp}{(N-mp+2)(mp-rp_1)} \\
& (tr\,[\,m\underline{G}\underline{C}_{22.1}^{-1}\,] + tr\,[\,(m-r)\underline{C}_{11}^{-1}\underline{X}_1^T\underline{Q}\underline{X}_1\,]),
\end{aligned} \tag{4.60}$$

where $\underline{G} = (\underline{X}_1\underline{C}_{11}^{-1}\underline{C}_{12} - \underline{X}_2)^T\underline{Q}\,(\underline{X}_1\underline{C}_{11}^{-1}\underline{C}_{12} - \underline{X}_2)$. From (4.60), it can be concluded that $r(\xi;\underline{e}_{1c},\cdots,\underline{e}_{mc}) < r(\xi;\underline{X}\tilde{\underline{\beta}}_1,\cdots,\underline{X}\tilde{\underline{\beta}}_m)$, when $0 < c < 2(mp - rp_1 - 2)$. Since $\underline{e}_{kc} = \underline{e}_{kEB}$ if $c = mp - rp_1 - 2$, Theorem 4.2.4 is proved.

The next theorem gives necessary and sufficient conditions under which the Bayes risk of $\underline{e}^*_{EB}$ is smaller than that of $\underline{e}^*_{RMLE}$.

Theorem 4.2.5. Consider the set of Theorem 4.2.4. Then,

$$r(\xi, \underline{e}_{1EB}, \cdots, \underline{e}_{mEB}) < r(\xi, \underline{X}_1\hat{\underline{D}}_1\underline{a}_1, \cdots, \underline{X}_1\hat{\underline{D}}_1\underline{a}_m), \tag{4.61}$$

if and only if $(1-B)^2/B^2 > 1-(mp-rp_1-2)(N-mp)/[(mp-rp_1)(N-mp+2)]$ with $B = \sigma^2/(\sigma^2+\tau^2)$.

Proof of Theorem 4.2.5. By expressing

$$\begin{aligned}
&\underline{X}_1\hat{\underline{D}}_1\underline{a}_k - \underline{X}\underline{\beta}_k = \underline{X}_1(\hat{\underline{D}}_1\underline{a}_k - E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}(\hat{\underline{D}}_1\underline{a}_k)) \\
+\ &\underline{X}_1(E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}(\hat{\underline{D}}_1\underline{a}_k) - E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}(\hat{\underline{\beta}}_{k1})) \\
+\ &(\underline{X}_1\underline{C}_{11}^{-1}\underline{C}_{12} - \underline{X}_2)\underline{\beta}_{k2},
\end{aligned}$$

it can be seen that

$$\begin{aligned}
&R(\underline{X}\underline{\beta}_1, \cdots, \underline{X}\underline{\beta}_m; \underline{X}_1\hat{\underline{D}}_1\underline{a}_1, \cdots, \underline{X}_1\hat{\underline{D}}_1\underline{a}_m) \\
=\ &\sum_{k=1}^{m}\{tr\,[\,V(\hat{\underline{D}}_1\underline{a}_k)(\underline{X}_1^T\underline{Q}\underline{X}_1)\,] \\
+\ &tr\,[\,E(\hat{\underline{D}}_1\underline{a}_k - \hat{\underline{\beta}}_{k1})E(\hat{\underline{D}}_1\underline{a}_k - \hat{\underline{\beta}}_{k1})^T(\underline{X}_1^T\underline{Q}\underline{X}_1)\,] \\
+\ &2tr\,[\,\underline{\beta}_{k2}E(\hat{\underline{D}}_1\underline{a}_k - \hat{\underline{\beta}}_{k1})^T(\underline{X}_1^T\underline{Q}(\underline{X}_1\underline{C}_{11}^{-1}\underline{C}_{12} - \underline{X}_2))\,] \\
+\ &tr\,[\underline{\beta}_{k2}\underline{\beta}_{k2}^T\underline{G}\,]\},
\end{aligned} \tag{4.62}$$

where $\underline{G} = (\underline{X}_1\underline{C}_{11}^{-1}\underline{C}_{12} - \underline{X}_2)^T\underline{Q}\ (\underline{X}_1\underline{C}_{11}^{-1}\underline{C}_{12} - \underline{X}_2)$. Since

$$\underline{\beta}_k \sim^{ind} N_p\Big(\begin{pmatrix} \underline{D}_1\underline{a}_k \\ \underline{0} \end{pmatrix}, \tau^2(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\Big), \tag{4.63}$$

then,

$$\begin{aligned}
&E_{\underline{D}_1\underline{a}_1,\cdots,\underline{D}_1\underline{a}_m,\tau^2}(\underline{\beta}_{k2}E_{\underline{\beta}_1,\cdots,\underline{\beta}_m}(\hat{\underline{D}}_1\underline{a}_k - \hat{\underline{\beta}}_{k1})^T) \\
=\ &Cov(E_{\underline{\beta}_1,\cdots,\underline{\beta}_m}(\hat{\underline{D}}_1\underline{a}_k - \hat{\underline{\beta}}_{k1}), \underline{\beta}_{k2})
\end{aligned}$$

$$
\begin{aligned}
&= Cov((\underline{\beta}_1^{\#T} + \underline{C}_{11}^{-1}\underline{C}_{12}\underline{\beta}_2^{\#T})(\underline{A}(\underline{A}^T\underline{A})^{-1}\underline{a}_k - \underline{1}_k), \underline{\beta}_2^{\#T}\underline{1}_k) \\
&= (-\underline{C}_{11}^{-1}\underline{C}_{12}\underline{C}_{22.1}^{-1} + \underline{C}_{11}^{-1}\underline{C}_{12}\underline{C}_{22.1}^{-1})(\underline{a}_k^T(\underline{A}^T\underline{A})^{-1}\underline{a}_k - 1) \\
&= \underline{0}, \qquad (4.64)
\end{aligned}
$$

where $\underline{\beta}_i^{\#} = (\underline{\beta}_{1i}, \cdots, \underline{\beta}_{mi})^T$, $i = 1, 2$ and $\underline{1}_k$ is the kth column of $\underline{I}_m$. And

$$
\begin{aligned}
&E_{\underline{D}_1\underline{a}_1, \cdots, \underline{D}_1\underline{a}_m, \tau^2}\{E(\underline{\hat{D}}_1\underline{a}_k - \underline{\hat{\beta}}_{k1})E(\underline{\hat{D}}_1\underline{a}_k - \underline{\hat{\beta}}_{k1})^T\} \\
&= V((\underline{\beta}_1^{\#T} - \underline{C}_{11}^{-1}\underline{C}_{12}\underline{\beta}_2^{\#T})(\underline{1}_k - \underline{A}(\underline{A}^T\underline{A})^{-1}\underline{a}_k)) \\
&= \tau^2(1 - \underline{a}_k^T(\underline{A}^T\underline{A})^{-1}\underline{a}_k)(\underline{C}_{11.2}^{-1} - \underline{C}_{11}^{-1}\underline{C}_{12}\underline{C}_{22.1}^{-1}\underline{C}_{21}\underline{C}_{11}^{-1}) \\
&= \tau^2(1 - \underline{a}_k^T(\underline{A}^T\underline{A})^{-1}\underline{a}_k)\underline{C}_{11}^{-1}. \qquad (4.65)
\end{aligned}
$$

By using (4.62),(4.63), (4.64), and (4.65), it can be obtained that

$$
\begin{aligned}
&r(\xi, \underline{X}_1\underline{\hat{D}}_1\underline{a}_1, \cdots, \underline{X}_1\underline{\hat{D}}_1\underline{a}_m) \\
&= \sigma^2 r\, tr[\,\underline{C}_{11}^{-1}(\underline{X}_1^T\underline{Q}\underline{X}_1)\,] \\
&+ \tau^2((m-r)tr[\,\underline{C}_{11}^{-1}\underline{X}_1^T\underline{Q}\underline{X}_1\,] \\
&+ m\, tr[\underline{C}_{22.1}^{-1}\underline{G}]. \qquad (4.66)
\end{aligned}
$$

Now,by using (4.60) with $c = mp - rp_1 - 2$ and (4.66),

$$
\begin{aligned}
&r(\xi, \underline{X}_1\underline{\hat{D}}_1\underline{a}_1, \cdots, \underline{X}_1\underline{\hat{D}}_1\underline{a}_m) - r(\xi, \underline{\varepsilon}_{1EB}, \cdots, \underline{\varepsilon}_{mEB}) \\
&= (\tau^2 - \sigma^2 + \sigma^2 B\frac{(N-mp)}{(N-mp+2)}\,\frac{(mp-rp_1-2)}{(mp-rp_1)}) \\
&\{(m-r)tr[\,\underline{C}_{11}^{-1}\underline{X}_1^T\underline{Q}\underline{X}_1\,] + m\,tr[\,\underline{C}_{22.1}^{-1}\underline{G}\,]\}. \qquad (4.67)
\end{aligned}
$$

Thus, the conclusion of the theorem follows from (4.67).

The consequence of the above theorem is clear. If $\sigma^2 < \tau^2$, then $B < 1/2$ in which case $(1-B)^2/B^2 > 1$. Therefore, $\underline{\varepsilon}_{EB}^*$ always dominates $\underline{\varepsilon}_{RMLE}^*$ in its Bayes risk if the sampling variability is smaller than the prior variability. However, if τ^2 is much smaller than σ^2, $\underline{\varepsilon}_{RMLE}^*$ becomes a more appropriate estimator of $(\underline{X}\underline{\beta})^*$.

A general class of minimax estimates of $(\underline{X}\underline{\beta})^*$, which was described earlier in this section, can be expressed as

$$\begin{aligned}
&\underline{e}_{kc}(\underline{Y}_k) \\
&= \underline{X}_{k1}(\hat{\underline{D}}_1\underline{a}_k + (1-\frac{c}{F})(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)) + \underline{X}_2(1-\frac{c}{F})\tilde{\underline{\beta}}_{k2},\ k=1,\cdots,m, \quad (4.68)
\end{aligned}$$

where c is a proper constant. The following theorem shows that $\underline{e}^*_{EB}$ is the optimal estimate of $(\underline{X}\underline{\beta})^*$ within the class of all estimators of the form $\underline{e}^*_c = (\underline{e}^T_{1c},\cdots,\underline{e}^T_{mc})^T$.

Theorem 4.2.6. Consider the same set up as of Theorem 4.2.4. Then,

$$\begin{aligned}
&r(\xi,\underline{e}_{1c},\cdots,\underline{e}_{mc}) - r(\xi,\underline{e}_{1B},\cdots,\underline{e}_{mB}) \\
&= \sigma^2 B m\, tr\,[(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\underline{X}^T\underline{Q}\underline{X}\,] \\
&+ (\frac{c^2}{mp-rp_1-2} - 2c)\sigma^2 B \frac{N-mp}{(N-mp+2)(mp-rp_1)} \\
&\quad \{m\, tr\,[\underline{G}\underline{C}^{-1}_{22.1}\,] + (m-r)tr\,[\underline{C}^{-1}_{11}(\underline{X}^T_1\underline{Q}\underline{X}_1)\,]\} \qquad (4.69)
\end{aligned}$$

For $mp - rp_1 \geq 3$, the above risk is minimized at $c = mp - rp_1 - 2$.

Proof of Theorem 4.2.6. From the proof of Theorem 2.2.8, it can be shown that

$$r(\xi,\underline{e}_{1B},\cdots,\underline{e}_{mB}) = \sigma^2(1-B)mtr\,[\,(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\underline{X}^T\underline{Q}\underline{X}\,] \qquad (4.70)$$

Then, Eq.(4.69) can be proved by combining (4.60) and (4.70).

4.3 Empirical Bayes Subset Estimators

In this section four empirical Bayes subset estimators of $\underline{\beta}_{k1}$ are discussed and compared under the loss in (3.69),

$$L(\underline{\beta}_{k1};\underline{a}_k) = (\underline{a}_k - \underline{\beta}_{k1})^T\underline{U}_k(\underline{a}_k - \underline{\beta}_{k1}).$$

The multivariate model is the same set as given in Theorem 4.2.4.

The shrinkage coefficients corresponding to the four empirical Bayes estimators will be derived from the usual F ratio for testing $H_0 : \underline{\beta}_{k2} = \underline{0}$. If $\underline{e}^1_{\phi k1}$ is defined as

an estimator of $\underline{\beta}_{k1}$ for some k that shrinks the unrestricted least squares estimator $\tilde{\underline{\beta}}_{k1}$ to the restricted least squares estimator $\hat{\underline{\beta}}_{k1}$ under the hypothesis $H_0 : \underline{\beta}_{k2} = \underline{0}$, then $\underline{e}^1_{\phi k1}$ can be expressed as

$$\underline{e}^1_{\phi k1} = \tilde{\underline{\beta}}_{k1} - \frac{\phi(F_k)}{F_k}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{k1}), \tag{4.71}$$

where $F_k = (N - mp + 2)\tilde{\underline{\beta}}^T_{k2}\underline{C}_{22.1}\tilde{\underline{\beta}}_{k2}/SSE$.

Now, $\underline{e}^2_{\phi k1}$ is defined as an estimator that shrinks the unrestricted least squares estimator $\tilde{\underline{\beta}}_{k1}$ to the restricted maximum likelihood estimator $\hat{\underline{D}}_1\underline{a}_k$ under the hypothesis $H_0 : \underline{\beta}_{k1} = \underline{D}_1\underline{a}_k$ and $\underline{\beta}_{k2} = \underline{0}$ for all k. If the shrinkage coefficient, $\phi(F_k)/F_k$, is the same as described in $\underline{e}^1_{k1}$, then

$$\underline{e}^2_{\phi k1} = \tilde{\underline{\beta}}_{k1} - \frac{\phi(F_k)}{F_k}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k). \tag{4.72}$$

Replacing the shrinkage coefficient $\phi(F_k)/F_k$ by $\phi(F)/F$ of $\underline{e}^1_{\phi k1}$ in (4.71) yields a third estimator as

$$\underline{e}^3_{\phi k1} = \tilde{\underline{\beta}}_{k1} - \frac{\phi(F)}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{\beta}}_{k1}), \tag{4.73}$$

where $F = (N - mp + 2)(SSH_1 + SSH_2)/SSE$ has been mentioned in Section 4.2.

The forth one is a subset of $\underline{e}_{k\phi}$ described in Section 4.2. Thus, $\underline{e}^4_{\phi k1}$ can be expressed as

$$\underline{e}^4_{\phi k1} = \tilde{\underline{\beta}}_{k1} - \frac{\phi(F)}{F}(\tilde{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k). \tag{4.74}$$

The comparison among these estimators are based on their frequentist risks and Bayes risks. Now, these estimators of $\underline{\beta}_{k1}$ are first evaluated for some individual k according to the Bayes criterion. Let $\phi(F) = \phi(F_k) = c$, a positive constant, and, as a convenience, rewrite $\underline{e}^i_{\phi k1}$ as $\underline{e}^i_{ck1}$ for $i = 1, 2, 3, 4$. The following theorem shows that under the prior described in Section 4.2 the Bayes risk of $\underline{e}^4_{ck1}$ with $c = mp - rp_1 - 2$, i.e., the subset of $\underline{e}_{kEB}$, is the smallest among the four empirical Bayes estimators of $\underline{\beta}_{k1}$ irrespective of any p.d. $\underline{U}_k$.

Theorem 4.3.1. Consider the prior ξ in Theorem 4.2.4. Let $r(\xi, \underline{e}_{k1})$ denote the Bayes risk of an estimator $\underline{e}_{k1}$ of $\underline{\beta}_{k1}$ for some individual k under the prior ξ when the loss is given in (3.69). Then,

$$\begin{aligned} & \min_c\{r(\xi, \underline{e}^4_{ck1})\} \\ = \ & \min\{\min_c\{r(\xi, \underline{e}^1_{ck1})\}, \min_c\{r(\xi, \underline{e}^2_{ck1})\}, \\ & \min_c\{r(\xi, \underline{e}^3_{ck1})\}, \min_c\{r(\xi, \underline{e}^4_{ck1})\}\}, \end{aligned}$$

$$\begin{aligned} & \min_c\{r(\xi, \underline{e}^1_{ck1})\} \\ = \ & \max\{\min_c\{r(\xi, \underline{e}^1_{ck1})\}, \min_c\{(\xi, \underline{e}^2_{ck1})\}, \\ & \min_c\{r(\xi, \underline{e}^3_{ck1})\}, \min_c\{(\xi, \underline{e}^4_{ck1})\}\}. \end{aligned}$$

Proof of Theorem 4.3.1. From the proof of Theorem 3.3.1, it can be obtained that

$$\begin{aligned} r(\xi, \underline{e}^1_{ck1}) &= \sigma^2 tr\,[\,\underline{C}^{-1}_{11.2}\underline{U}_k\,] + \sigma^2 B \frac{N - mp}{(N - mp + 2)p_2}\left(\frac{c^2}{p_2 - 2} - 2c\right) \\ &\times \ tr\,[\,(\underline{C}^{-1}_{11.2} - \underline{C}^{-1}_{11})\underline{U}_k\,], \end{aligned} \tag{4.75}$$

which is minimized at $c = p_2 - 2$. The minimum of $r(\xi, \underline{e}^1_{ck1})$ is

$$\begin{aligned} r(\xi, \underline{e}^1_{(p_2-2)k1}) &= \sigma^2 tr\,[\,\underline{C}^{-1}_{11.2}\underline{U}_k\,] \\ &- \sigma^2 B \frac{(N - mp)(p_2 - 2)}{(N - mp + 2)p_2} tr\,[\,(\underline{C}^{-1}_{11.2} - \underline{C}^{-1}_{11})\underline{U}_k\,]. \end{aligned} \tag{4.76}$$

By following the proof of Theorem 3.3.1, using $\underline{\hat{D}}_1\underline{a}_k$ instead of $\underline{\hat{\beta}}_{.1}$, and considering $E(\underline{\hat{\beta}}_{k1} - \underline{\hat{D}}_1\underline{a}_k)(\underline{\hat{\beta}}_{k1} - \underline{\hat{D}}_1\underline{a}_k)^T = V(\underline{\hat{\beta}}_{k1} - \underline{\hat{D}}_1\underline{a}_k) = (\sigma^2 + \tau^2)(1 - \underline{a}^T_k(\underline{A}^T\underline{A})^{-1}\underline{a}_k)\underline{C}^{-1}_{11}$ and the independence of $\underline{\hat{D}}_1\underline{a}_k$, and $\underline{\hat{\beta}}_{k1} - \underline{\hat{D}}_1\underline{a}_k$, it can be shown that

$$\begin{aligned} r(\xi, \underline{e}^2_{ck1}) &= \sigma^2 tr\,[\,\underline{C}^{-1}_{11.2}\underline{U}_k\,] + \sigma^2 B \frac{N - mp}{(N - mp + 2)p_2}\left(\frac{c^2}{p_2 - 2} - 2c\right) \\ &\times \ tr\,[\,((1 - \underline{a}^T_k(\underline{A}^T\underline{A})^{-1}\underline{a}_k)\underline{C}^{-1}_{11} + (\underline{C}^{-1}_{11.2} - \underline{C}^{-1}_{11}))\underline{U}_k\,], \end{aligned} \tag{4.77}$$

which is minimized at $c = p_2 - 2$. Thus, the minimum of $r(\xi, \underline{e}^2_{c\,k1})$ is

$$\begin{aligned} r(\xi, \underline{e}^2_{(p_2-2)k1}) &= \sigma^2 tr\,[\,\underline{C}^{-1}_{11.2}\underline{U}_k\,] - \sigma^2 B \frac{(N-mp)(p_2-2)}{(N-mp+2)p_2} \\ &\times\ tr\,[\,((1-\underline{a}_k(\underline{A}^T\underline{A})^{-1}\underline{a}_k)\underline{C}^{-1}_{11} + (\underline{C}^{-1}_{11.2} - \underline{C}^{-1}_{11}))\underline{U}_k\,]. \quad (4.78) \end{aligned}$$

The Bayes risk of $\underline{e}^3_{c\,k1}$ also can be expressed as

$$\begin{aligned} r(\xi, \underline{e}^3_{c\,k1}) &= \sigma^2 tr\,[\,\underline{C}^{-1}_{11.2}\underline{U}_k\,] \\ &+\ \sigma^2 B \frac{N-mp}{(N-mp+2)(mp-rp_1)}\left(\frac{c^2}{mp-rp_1-2} - 2c\right) \\ &\times\ tr\,[\,(\underline{C}^{-1}_{11.2} - \underline{C}^{-1}_{11})\underline{U}_k\,], \end{aligned} \quad (4.79)$$

which is minimized at $c = mp - rp_1 - 2$. Thus, the minimum of $r(\xi, \underline{e}^3_{c\,k1})$ becomes

$$\begin{aligned} r(\xi, \underline{e}^2_{(mp-rp_1-2)k1}) &= \sigma^2 tr\,[\,\underline{C}^{-1}_{11.2}\underline{U}_k\,] - \sigma^2 B \frac{(N-mp)(mp-rp_1-2)}{(N-mp+2)(mp-rp_1)} \\ &\times\ tr\,[\,(\underline{C}^{-1}_{11.2} - \underline{C}^{-1}_{11})\underline{U}_k\,]. \end{aligned} \quad (4.80)$$

By using (4.55),

$$\begin{aligned} r(\xi, \underline{e}^4_{c\,k1}) &= \sigma^2 tr\,[\,\underline{C}^{-1}_{11.2}\underline{U}_k\,] \\ &+\ \sigma^2 B \frac{N-mp}{(N-mp+2)(mp-rp_1)}\left(\frac{c^2}{mp-rp_1-2} - 2c\right) \\ &\times\ tr\,[\,((1-\underline{a}^T_k(\underline{A}^T\underline{A})^{-1}\underline{a}_k)\underline{C}^{-1}_{11} + (\underline{C}^{-1}_{11.2} - \underline{C}^{-1}_{11}))\underline{U}_k\,], \end{aligned} \quad (4.81)$$

which is minimized at $c = mp - rp_1 - 2$. Thus, the minimum of $r(\xi, \underline{e}^4_{c\,k1})$ is

$$\begin{aligned} r(\xi, \underline{e}^2_{(mp-rp_1-2)k1}) &= \sigma^2 tr\,[\,\underline{C}^{-1}_{11.2}\underline{U}_k\,] \\ &-\ \sigma^2 B \frac{(N-mp)(mp-rp_1-2)}{(N-mp+2)(mp-rp_1)} \\ &\times\ tr\,[\,(1-\underline{a}^T_k(\underline{A}^T\underline{A})^{-1}\underline{a}_k)\underline{C}^{-1}_{11} + (\underline{C}^{-1}_{11.2} - \underline{C}^{-1}_{11}))\underline{U}_k\,]. \end{aligned} \quad (4.82)$$

Comparing (4.76), (4.78), (4.80) and (4.82), Theorem 4.3.1 follows.

The condition described in Theorem 4.3.1 is clearly expected. This is because $\underline{e}^4_{c\,k1}$ use all of $\underline{Y}_1, \cdots, \underline{Y}_m$ in estimating $\underline{D}_1$ as well as $1/(\sigma^2+\tau^2)$ while the other three estimators use only $\underline{Y}_k$ but not all of $\underline{Y}_1, \cdots, \underline{Y}_m$ for estimating either $\underline{D}_1$ or $1/(\sigma^2+\tau^2)$ or both.

If c is assumed to satisfy the conditions of that $0 < c < 2(p_2 - 2)$ in $\underline{e}^1_{ck1}$ and $\underline{e}^2_{ck1}$, and $0 < c < 2(mp - rp_1 - 2)$ in $\underline{e}^3_{ck1}$ and $\underline{e}^4_{ck1}$, it is then easy to demonstrate that all these four estimators of $\underline{\beta}_{k1}$ for some k is better than $\tilde{\underline{\beta}}_{k1}$.

Suppose one expresses

$$\begin{aligned}\hat{\underline{\beta}}_{k1} - \underline{\beta}_{k1} &= (\hat{\underline{\beta}}_{k1} - E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}(\hat{\underline{\beta}}_{k1})) \\ &+ \underline{C}^{-1}_{11}\underline{C}_{12}\underline{\beta}_{k2},\end{aligned}$$

and gets

$$\begin{aligned}& r(\xi, \hat{\underline{\beta}}_{k1}) \\ &= \sigma^2 tr(\underline{C}^{-1}_{11}\underline{U}_k) + \tau^2 tr\,[\,(\underline{C}^{-1}_{11.2} - \underline{C}^{-1}_{11})\underline{U}_k). \qquad (4.83)\end{aligned}$$

Therefore,

$$\begin{aligned}& r(\xi, \hat{\underline{\beta}}_{k1}) - r(\xi, \underline{e}^1_{(p_2-2)k1}) \\ &= (\tau^2 - \sigma^2 + \sigma^2 B\frac{(N-mp)(p_2-2)}{(N-mp+2)p_2})tr\,[\,(\underline{C}^{-1}_{11.2} - \underline{C}^{-1}_{11})\underline{U}_k\,] > 0, \qquad (4.84)\end{aligned}$$

if and only if $(1-B)^2/B^2 > 1 - (N-mp)(p_2-2)/((N-mp+2)p_2)$, and

$$\begin{aligned}& r(\xi, \hat{\underline{\beta}}_{k1}) - r(\xi, \underline{e}^2_{(p_2-2)k1}) \\ &= \sigma^2 B\frac{(N-mp)(p_2-2)}{(N-mp+2)p_2})tr\,[\,(1 - \underline{a}^T_k(\underline{A}^T\underline{A})^{-1}\underline{a}_k)\underline{C}^{-1}_{11}\underline{U}_k\,] \\ &+ (\tau^2 - \sigma^2 + \sigma^2 B\frac{(N-mp)(p_2-2)}{(N-mp+2)p_2})tr\,[\,(\underline{C}^{-1}_{11.2} - \underline{C}^{-1}_{11})\underline{U}_k\,] > 0, \qquad (4.85)\end{aligned}$$

if and only if $(1-B)^2/B^2 > 1 - (N-mp)(p_2-2)/((N-mp+2)p_2)$.

Also based on the same procedure given above,

$$\begin{aligned}& r(\xi, \hat{\underline{\beta}}_{k1}) - r(\xi, \underline{e}^3_{(mp-rp_1-2)k1}) \\ &= (\tau^2 - \sigma^2 + \sigma^2 B\frac{(N-mp)(mp-rp_1-2)}{(N-mp+2)(mp-rp_1)}tr((\underline{C}^{-1}_{11.2} - \underline{C}^{-1}_{11})\underline{U}_k) > 0, \qquad (4.86)\end{aligned}$$

if and only if $(1-B)^2/B^2 > 1 - (N-mp)(mp-rp_1-2)/((N-mp+2)(mp-rp_1))$, and

$$
\begin{aligned}
& r(\xi, \hat{\underline{\beta}}_{k1}) - r(\xi, \underline{e}^4_{(mp-rp_1-2)k1}) \\
= \; & B\frac{(N-mp)(mp-rp_1-2)}{(N-mp+2)(mp-rp_1)} tr\,[\,(1-\underline{a}_k^T(\underline{A}^T\underline{A})^{-1}\underline{a}_k)\underline{C}_{11.2}^{-1}\underline{U}_k\,] \\
+ \; & (\tau^2-\sigma^2+\sigma^2 B\frac{(N-mp)(mp-rp_1-2)}{(N-mp+2)(mp-rp_1)}) tr\,[\,(\underline{C}_{11.2}^{-1}-\underline{C}_{11}^{-1})\underline{U}_k\,] > 0, \quad (4.87)
\end{aligned}
$$

if and only if $(1-B)^2/B^2 > 1-(N-mp)(mp-rp_1-2)/((N-mp+2)(mp-rp_1))$.

The results of (4.84), (4.85), (4.86), and (4.87) are not surprising since $\hat{\underline{\beta}}_{k1}$ is expected to perform better than $\underline{e}^i_{ck1}$'s (i=1,2,3,4) when the prior variance τ^2 is much smaller than the sampling variance σ^2. However, all these four $\underline{e}^i_{ck1}$'s can have smaller Bayes risks than $\hat{\underline{\beta}}_{k1}$ if $\tau^2 > \sigma^2$.

Next, the frequentist criterion is utilized to evaluate the four estimators. From Eq. (3.94), it can be shown that

$$
\begin{aligned}
& R(\underline{\beta}_{k1}, \underline{e}^1_{\phi k1}) \\
= \; & \sigma^2 tr\,[\,\underline{C}_{11.2}^{-1}\underline{U}_k\,] - 2\sigma^2 E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}(2(\frac{\phi'(F_k)}{F_k} - \frac{\phi(F_k)}{F_k^2}) \\
\times \; & \tilde{\underline{\beta}}_{k2}^T\underline{C}_{21}\underline{C}_{11}^{-1}\underline{U}_k\underline{C}_{11}^{-1}\underline{C}_{12}\tilde{\underline{\beta}}_{k2}/(SSE/(N-mp+2)) \\
+ \; & \frac{\phi(F_k)}{F_k} tr\,[(\underline{C}_{11.2}^{-1}-\underline{C}_{11}^{-1})\underline{U}_k]) \\
+ \; & E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{\frac{\phi^2(F_k)}{F_k^2}(\tilde{\underline{\beta}}_{k2}^T\underline{C}_{21}\underline{C}_{11}^{-1}\underline{U}_k\underline{C}_{11}^{-1}\underline{C}_{12}\tilde{\underline{\beta}}_{k2})\}. \qquad (4.88)
\end{aligned}
$$

The frequentist risk of $\underline{e}^2_{\phi k1}$ can be then obtained as

$$
\begin{aligned}
& R(\underline{\beta}_{k1}, \underline{e}^2_{\phi k1}) \\
= \; & E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2} tr\,[\,(\hat{\underline{\beta}}_{k1}-\underline{\beta}_{k1})(\hat{\underline{\beta}}_{k1}-\underline{\beta}_{k1})^T\underline{U}_k\,] \\
- \; & 2tr\,[\,E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{((\hat{\underline{\beta}}_{k1}-E(\hat{\underline{\beta}}_{k1})) - \underline{C}_{11}^{-1}\underline{C}_{12}(\tilde{\underline{\beta}}_{k2}-\underline{\beta}_{k2})) \\
\times \; & \frac{\phi(F_k)}{F_k}(\hat{\underline{\beta}}_{k1}-\underline{\hat{D}}_1\underline{a}_k - \underline{C}_{11}^{-1}\underline{C}_{12}\tilde{\underline{\beta}}_{k2})^T\}\underline{U}_k\,] \\
+ \; & E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{\frac{\phi^2(F_k)}{F_k^2}((\hat{\underline{\beta}}_{k1}-\underline{\hat{D}}_1\underline{a}_k)^T\underline{U}_k(\hat{\underline{\beta}}_{k1}-\underline{\hat{D}}_1\underline{a}_k) \\
- \; & 2(\underline{C}_{11}^{-1}\underline{C}_{12}\tilde{\underline{\beta}}_{k2})^T\underline{U}_k(\hat{\underline{\beta}}_{k1}-\underline{\hat{D}}_1\underline{a}_k) + \tilde{\underline{\beta}}_{k2}^T\underline{C}_{21}\underline{C}_{11}^{-1}\underline{U}_k\underline{C}_{11}^{-1}\underline{C}_{12}\tilde{\underline{\beta}}_{k2})\}. \qquad (4.89)
\end{aligned}
$$

By using a generalization of Stein's identity (Berger and Haff, 1983), the cross-product term of (4.89) can be expressed as

$$
\begin{aligned}
& 2\sigma^2 tr\,[\, E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{2(\frac{\phi'(F_k)}{F_k} - \frac{\phi(F_k)}{F_k^2})(\underline{C}_{11}^{-1}\underline{C}_{12}\tilde{\underline{\beta}}_{k2}\tilde{\underline{\beta}}_{k2}^T\underline{C}_{21}\underline{C}_{11}^{-1} \\
- \; & \underline{C}_{11}^{-1}\underline{C}_{12}\tilde{\underline{\beta}}_{k2}(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T)/(SSE/(N-mp+2)) \\
+ \; & \frac{\phi(F_k)}{F_k}(\underline{C}_{11.2}^{-1} - (\underline{a}_k^T(\underline{A}^T\underline{A})^{-1}\underline{a}_k)\underline{C}_{11}^{-1})\}\underline{U}_k\,]. \qquad (4.90)
\end{aligned}
$$

From (4.89) and (4.90),

$$
\begin{aligned}
& R(\underline{\beta}_{k1}, \underline{e}_{\phi k1}^2) \\
= \; & \sigma^2 tr(\underline{C}_{11.2}^{-1}\underline{U}_k) - 2\sigma^2 E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{2(\frac{\phi'(F_k)}{F_k} - \frac{\phi(F_k)}{F_k^2}) \\
\times \; & (\tilde{\underline{\beta}}_{k2}^T\underline{C}_{21}\underline{C}_{11}^{-1}\underline{U}_k\underline{C}_{11}^{-1}\underline{C}_{12}\tilde{\underline{\beta}}_{k2} - 2(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T\underline{U}_k\underline{C}_{11}^{-1}\underline{C}_{12}\tilde{\underline{\beta}}_{k2}) \\
\div \; & (SSE/(N-mp+2)) \\
+ \; & \frac{\phi(F_k)}{F_k} tr\,[\,(\underline{C}_{11.2}^{-1} - (\underline{a}_k(\underline{A}^T\underline{A})^{-1}\underline{a}_k)\underline{C}_{11}^{-1})\underline{U}_k\,]\} \\
+ \; & E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{\frac{\phi^2(F_k)}{F_k^2}((\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T\underline{U}_k(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k) \\
- \; & 2(\underline{C}_{11}^{-1}\underline{C}_{12}\tilde{\underline{\beta}}_{k2})^T\underline{U}_k(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k) + \tilde{\underline{\beta}}_{k2}^T\underline{C}_{21}\underline{C}_{11}^{-1}\underline{U}_k\underline{C}_{11}^{-1}\underline{C}_{12}\tilde{\underline{\beta}}_{k2})\}. \qquad (4.91)
\end{aligned}
$$

By using the same argument, it can be shown that

$$
\begin{aligned}
& R(\underline{\beta}_{k1}, \underline{e}_{\phi k1}^3) \\
= \; & \sigma^2 tr(\underline{C}_{11.2}^{-1}\underline{U}_k) - 2\sigma^2 E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{2(\frac{\phi'(F)}{F} - \frac{\phi(F)}{F^2}) \\
\times \; & (\tilde{\underline{\beta}}_{k2}^T\underline{C}_{21}\underline{C}_{11}^{-1}\underline{U}_k\underline{C}_{11}^{-1}\underline{C}_{12}\tilde{\underline{\beta}}_{k2} - (\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T\underline{U}_k\underline{C}_{11}^{-1}\underline{C}_{12}\tilde{\underline{\beta}}_{k2}) \\
\div \; & (SSE/(N-mp+2)) + \frac{\phi(F)}{F} tr((\underline{C}_{11.2}^{-1} - \underline{C}_{11}^{-1})\underline{U}_k)\} \\
+ \; & E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}(\frac{\phi^2(F)}{F^2}\tilde{\underline{\beta}}_{k2}^T\underline{C}_{21}\underline{C}_{11}^{-1}\underline{U}_k\underline{C}_{11}^{-1}\underline{C}_{12}\tilde{\underline{\beta}}_{k2}), \qquad (4.92)
\end{aligned}
$$

and

$$
\begin{aligned}
& R(\underline{\beta}_{k1}, \underline{e}_{\phi k1}^4) \\
= \; & \sigma^2 tr(\underline{C}_{11.2}^{-1}\underline{U}_k) - 2\sigma^2 E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{2(\frac{\phi'(F)}{F} - \frac{\phi(F)}{F^2})
\end{aligned}
$$

$$
\begin{aligned}
&((\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T\underline{U}_k(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k) - 2(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T\underline{U}_k\underline{C}_{11}^{-1}\underline{C}_{12}\tilde{\underline{\beta}}_{k2} \\
&+ \tilde{\underline{\beta}}_{k2}^T\underline{C}_{21}\underline{C}_{11}^{-1}\underline{U}_k\underline{C}_{11}^{-1}\underline{C}_{12}\tilde{\underline{\beta}}_{k2})/(SSE/(N-mp+2)) \\
&+ \frac{\phi(F)}{F}tr\,[\,(\underline{C}_{11.2}^{-1} - (\underline{a}_k^T(\underline{A}^T\underline{A})^{-1}\underline{a}_k)\underline{C}_{11}^{-1})\underline{U}_k\,]\} \\
&+ E_{\underline{\beta}_1,\cdots,\underline{\beta}_m,\sigma^2}\{\frac{\phi^2(F)}{F^2}((\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T\underline{U}_k(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k) \\
&- 2(\hat{\underline{\beta}}_{k1} - \hat{\underline{D}}_1\underline{a}_k)^T\underline{U}_k\underline{C}_{11}^{-1}\underline{C}_{12}\tilde{\underline{\beta}}_{k2} + \tilde{\underline{\beta}}_{k2}^T\underline{C}_{21}\underline{C}_{11}^{-1}\underline{U}_k\underline{C}_{11}^{-1}\underline{C}_{12}\tilde{\underline{\beta}}_{k2}\}. \qquad (4.93)
\end{aligned}
$$

Investigations of Eqs.(4.88), (4.91), (4.92) and (4.93) indicate that there is no evident winner among these four estimators as demonstrated by the comparisons of their frequentist risk with the loss defined in (3.69). This remains true even though $\phi(F) = \phi(F_k)$ is equal to a positive constant c.

4.4 Numerical Examples

To illustrate the behavior of the estimators derived in the previous sections, a Monte-Carlo experiment was conducted to assess the frequentist risks as well as the Bayes risks of the different estimators.

The same experiment described in Section 3.4 is used here. Random samples of $m = 3$ normally distributed random vectors $\underline{Y}_k$ $(k = 1, 2, 3)$ of dimension 16×1 are taken in a 2^3 factorial experiment. Here $\underline{Y}_k$ is the vector response of the *kth* individual, which is based on a 2^3 factorial experiment replicated two times (see Section 2.4). It is assumed that

$$
\begin{aligned}
E(Y_{ijk}) &= \gamma_{0k} + \gamma_{1k}\zeta_{1i} + \gamma_{2k}\zeta_{2i} + \gamma_{3k}\zeta_{3i} \\
&+ \gamma_{12k}\zeta_{1i}\zeta_{2i} + \gamma_{13k}\zeta_{1i}\zeta_{3i} + \gamma_{23k}\zeta_{2i}\zeta_{3i} + \gamma_{123k}\zeta_{1i}\zeta_{2i}\zeta_{3i},
\end{aligned}
$$

where $i = 1, \cdots, 8$, $j = 1, 2$, and $k = 1, 2, 3$. Recall that $\underline{\beta}_{k1}^T = (\gamma_{0k}, \gamma_{1k}, \gamma_{2k}, \gamma_{3k})$ and $\underline{\beta}_{k2}^T = (\gamma_{12k}, \gamma_{13k}, \gamma_{23k}, \gamma_{123k})$, $mp_1 = mp_2 = 12$ and $N = 48$. The design matrix $\underline{X}$ can be partitioned as $\underline{X} = (\underline{X}_1, \underline{X}_2)$, for all k, where

$$
\underline{X}_1^T = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\
1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1
\end{bmatrix},
$$

$$\underline{X}_2^T = \begin{bmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \end{bmatrix}.$$

It is assumed here that $\underline{V}(\underline{Y}_k) = \underline{I}_{16}$, for all k.

First, the simulated frequentist risks was evaluated based on the loss in (4.3) with $\underline{Q}_k = \underline{I}_{16}$ for all k. Random vectors $\underline{Y}_k$'s are generated from $N_{16}(\underline{X}\underline{\beta}_k, \underline{I}_{16})$ in each replication. The sampling procedure is repeated 1,000 times in each simulation. The estimated frequentist risks of estimating $\underline{X}\underline{\beta}_k$ $(k = 1, 2, 3)$, i.e., the average squared error losses of the different estimators are compared.

The choice of the $\underline{\beta}_k$'s is motivated from a null hypothesis $H_0 : \underline{\beta}_k = \underline{D}\underline{a}_k$ and departures from this null hypothesis. Accordingly, these $\underline{\beta}_k$'s are of the form $\underline{\beta}_k = \underline{D}\underline{a}_k + \underline{r}_k$ (the $\underline{r}_k$'s denoting the deviations from the null hypothesis for each k).

The matrix $\underline{D}$ is chosen as

$$\underline{D}^T = (\underline{D}_1^T, \underline{D}_2^T) = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & -2 & 3 & -2 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and

$$\underline{a}_1 = \begin{bmatrix} -1 \\ 3 \end{bmatrix}, \ \underline{a}_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \text{ and } \underline{a}_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

There are several choices of the $\underline{r}_k$'s. In the first case, $\underline{r}_k = \underline{0}$ for all k; in the second case that $\underline{r}_1 = \underline{r}_3 = \underline{0}$ and $\underline{r}_2 = (10\ -110000)^T$, in the third case that $\underline{r}_1 = \underline{r}_3 = \underline{0}$ and $\underline{r}_2 = (10\ -111234)^T$, and $\underline{r}_1 = \underline{0}, \underline{r}_2 = (10\ -111234)^T$ and $\underline{r}_3 = (00001111)^T$ in the last case. The results of the simulated frequentist risks are shown in Table 4.1.

In the first case, $\underline{\beta}_{k1} = \underline{D}_1\underline{a}_k$ and $\underline{\beta}_{k2} = \underline{0}$ for all k, the weighted average of LSE's for the reduced model, $\underline{\varepsilon}^*_{RMLE} = ((\underline{X}_1\hat{\underline{D}}_1\underline{a}_1)^T, (\underline{X}_1\hat{\underline{D}}_1\underline{a}_2)^T, (\underline{X}_1\hat{\underline{D}}_1\underline{a}_3)^T)^T$ is the best estimator of $(\underline{X}\underline{\beta})^*$. However, in the second case, which has $\underline{r}_2 \neq \underline{0}$, $\underline{\varepsilon}^*_{RLSE} = ((\underline{X}_1\hat{\underline{\beta}}_{11})^T, (\underline{X}_1\hat{\underline{\beta}}_{21})^T, (\underline{X}_1\hat{\underline{\beta}}_{31})^T)^T$ is the best as an estimator of $(\underline{X}\underline{\beta})^*$.

Table 4.1: The estimated frequentist risk results for 2^3 factorial experiment.

The	$\underline{\beta}_1$	6 −7 8 −6 0000	6 −7 8 −6 0000	6 −7 8 −6 0000	6 −7 8 −6 0000
esti.of	$\underline{\beta}_2$	4 −5 5 −4 0000	5 −5 4 −3 0000	5 −5 4 −3 1234	5 −5 4 −3 1234
$\underline{X}\underline{\beta}_k$'s	$\underline{\beta}_3$	2 −1 4 −2 0000	2 −1 4 −2 0000	2 −1 4 −2 0000	2 −1 4 −2 1111
$\underline{e}^*_{LSE}$		23.7611	23.7611	23.7611	23.7611
$\underline{e}^*_{RMLE}$		7.94026	37.4787	517.479	581.479
$\underline{e}^*_{RLSE}$		11.9607	11.9607	491.961	555.961
$\underline{e}^*_{EB}$		15.1057	20.9210	23.5225	23.5418
$\underline{e}^{+*}_{EB}$		14.9473	20.9210	23.5225	23.5418
$\underline{e}^*_{PTE.10}$		17.6193	23.8398	23.7611	23.7611
$\underline{e}^*_{MEB.10}$		13.4669	20.9913	23.5225	23.5418
$\underline{e}^{+*}_{MEB.10}$		13.4669	20.9913	23.5225	23.5418
$\underline{e}^*_{PTE.05}$		14.5875	24.0688	23.7611	23.7611
$\underline{e}^*_{MEB.05}$		12.0763	21.2337	23.5225	23.5418
$\underline{e}^{+*}_{MEB.05}$		12.0763	21.2337	23.5225	23.5418
$\underline{e}^*_{PTE.01}$		10.5772	25.3180	23.7611	23.7611
$\underline{e}^*_{MEB.01}$		9.79580	22.6270	23.5225	23.5418
$\underline{e}^{+*}_{MEB.01}$		9.79580	22.6270	23.5225	23.5418

* Note: For the right two columns, during the 1,000 replications, the null hypothesis, $\underline{\beta}_{k1} = \underline{D}_1\underline{a}_k$ and $\underline{\beta}_{k2} = \underline{0}$ $(k = 1, 2, 3)$, was rejected in every replication for all α =.10, .05 and .01. Hence, $\underline{e}_{kPTE\alpha} = \underline{X}\underline{\tilde{\beta}}_k$, $\underline{e}_{kMEB\alpha} = \underline{e}_{kEB}$ and $\underline{e}^+_{kMEB\alpha} = \underline{e}^+_{kEB}$.

Substantial improvement in the estimated frequentist risk of $\underline{\varepsilon}^{+*}_{EB}$ over $\underline{\varepsilon}^{*}_{LSE} = ((\underline{X}\tilde{\underline{\beta}}_1)^T, (\underline{X}\tilde{\underline{\beta}}_2)^T, (\underline{X}\tilde{\underline{\beta}}_3)^T)^T$, the LSE of the model in (4.1), was observed in the first two cases for which $\underline{\beta}_{k2} = \underline{0}$ $(k = 1, 2, 3)$. Especially in the first case, the estimated frequentist risk of $\underline{\varepsilon}^{+*}_{EB}$ is found only about one half of magnitude as that of $\underline{\varepsilon}^{*}_{LSE}$. As proved in the earlier section, the estimators $\underline{\varepsilon}^{*}_{MEB}$ or $\underline{\varepsilon}^{+*}_{MEB}$ performed better than $\underline{\varepsilon}^{*}_{PTE}$. Finally, if $\underline{\beta}_{k2} \neq \underline{0}$ for some k e.g. the third and the forth cases, $\underline{\varepsilon}^{*}_{EB}$ or $\underline{\varepsilon}^{+*}_{EB}$ has the smallest frequentist risks. Admittedly, the risk improvement is not significant over $\underline{\varepsilon}^{*}_{LSE}$, the LSE of the model in (4.1). All other estimators are superior over $\underline{\varepsilon}^{*}_{RMLE}$ and $\underline{\varepsilon}^{*}_{RLSE}$ with considerably less estimated frequentist risks.

Table 4.2 shows the estimated Bayes risks of all the estimators. First, $\underline{\beta}_k$ $(k = 1, 2, 3)$ are generated separately from $N(\underline{D}\underline{a}_k, \tau^2(\underline{X}^T\underline{X})^{-1})$ where

$$\underline{A}^T = (\underline{a}_1, \underline{a}_2, \underline{a}_3) = \begin{bmatrix} -1 & -1 & 1 \\ 3 & 2 & 1 \end{bmatrix},$$

and

$$\underline{D}^T = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & -2 & 3 & -2 & 0 & 0 & 0 & 0 \end{bmatrix},$$

or

$$\underline{D}^T = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 2 & 0 & 1 \\ 2 & -2 & 3 & -2 & 1 & 0 & 2 & 0 \end{bmatrix}.$$

For a generated set of $\underline{\beta}_k$'s and given σ^2, random vectors $\underline{Y}_k$ $(k = 1, 2, 3)$ are then generated separately from $N(\underline{X}\underline{\beta}_k, \sigma^2\underline{I}_{16})$. Once again, the sampling procedure is repeated 1,000 times in each simulation.

When $\tau^2 = .25$ and $\sigma^2 = 1$, i.e., when the prior variability is smaller than the sampling variability, the estimated Bayes risks of the different estimators maintain almost the same order as the corresponding estimated frequentist risks. This is reflected very clearly when $\underline{D}_2 = \underline{0}$, and to a certain extent when $\underline{D}_2 \neq \underline{0}$. Compare, for example, column 2 of Table 4.2 with column 2 of Table 4.1 or column 5 of Table 4.2 with Columns 4 and 5 of Table 4.1. However, the situation needs not be

Table 4.2: The estimated Bayes risk results for 2^3 factorial experiment.

	$\underline{Da}_1$	$(6\ -7\ 8\ -6\ 0\ 0\ 0)$			$(6\ -7\ 8\ -6\ 3\ -2\ 6\ -1)$		
	$\underline{Da}_2$	$(4\ -5\ 5\ -4\ 0\ 0\ 0)$			$(4\ -5\ 5\ -4\ 2\ -2\ 4\ -1)$		
The	$\underline{Da}_3$	$(2\ -1\ 4\ -2\ 0\ 0\ 0)$			$(2\ -1\ 4\ -2\ 1\ 2\ 2\ 1)$		
estimator	σ^2	1			1		
of $\underline{X}\underline{\beta}_k$'s	τ^2	.25	1	4	.25	1	4
$\underline{e}^*_{LSE}$		23.8374	23.8374	23.8374	23.8374	23.8374	23.8374
$\underline{e}^*_{RMLE}$		11.8254	23.8822	72.1095	1371.96	1384.15	1432.64
$\underline{e}^*_{RLSE}$		14.8925	23.9566	60.2131	1375.02	1384.22	1420.74
$\underline{e}^*_{EB}$		16.8208	19.4551	22.1019	23.7426	23.7535	23.7467
$\underline{e}^{+*}_{EB}$		16.7714	19.4551	22.1019	23.7426	23.7435	23.7467
$\underline{e}^*_{PTE.10}$		21.3026	23.8001	23.8374	23.8374	23.8374	23.8374
$\underline{e}^*_{MEB.10}$		16.4768	19.6594	22.1019	23.7426	23.7435	23.7467
$\underline{e}^{+*}_{MEB.10}$		16.4768	19.6594	22.1019	23.7426	23.7435	23.7467
$\underline{e}^*_{PTE.05}$		19.4267	23.7546	23.8464	23.8374	23.8374	23.8374
$\underline{e}^*_{MEB.05}$		15.8970	19.9896	22.1129	23.7426	23.7435	23.7467
$\underline{e}^{+*}_{MEB.05}$		15.8970	19.9896	22.1129	23.7426	23.7435	23.7467
$\underline{e}^*_{PTE.01}$		15.8870	23.9405	24.1212	23.8374	23.8374	23.8374
$\underline{e}^*_{MEB.01}$		14.1719	21.2385	22.3997	23.7426	23.7435	23.7467
$\underline{e}^{+*}_{MEB.01}$		14.1719	21.2385	22.3997	23.7426	23.7435	23.7467

* Note: For the right three columns, during the 1,000 replications, the null hypothesis, $\underline{\beta}_{k1} = \underline{D}_1\underline{a}_k$ and $\underline{\beta}_{k2} = \underline{0}$ $(k = 1, 2, 3)$, was rejected in every replication for all α =.10, .05 and .01. Hence, $\underline{e}_{kPTE\alpha} = \underline{X}\tilde{\underline{\beta}}_k$, $\underline{e}_{kMEB\alpha} = \underline{e}_{kEB}$ and $\underline{e}^+_{kMEB\alpha} = \underline{e}^+_{kEB}$.

the same when the prior variance equal to or greater than the sampling variance when $\underline{D} = \underline{0}$. For example while $\sigma^2 = 1$ and $\tau^2 = 4$, $\underline{e}^*_{RMLE}$ performs poorly when compared to the other estimators. The increased prior variability diminishes the importance of the prior mean. However, when $\underline{D}_2 \neq \underline{0}$, the estimators $\underline{e}^*_{RMLE}$ and $\underline{e}^*_{RLSE}$ perform poorly as anticiapted. The relative magnitude of the prior variances seems to play a less important role there. But the empirical Bayes estimators continue to maintain their superiority over the $\underline{e}^*_{LSE}$. As expected, this domination is more heavily pronounced when $\underline{D}_2 = \underline{0}$, and the prior variability is small when compared to the sampling variability.

CHAPTER 5
CONCLUSIONS

5.1 Summary and Conclusions

This study provides methods for estimating the response function and multivariate regression models where there is a choice between a full model and a reduced model. The classical approach is to test a hypothesis. If the null hypothesis is rejected at a desired level of significance the full model is used, otherwise, the reduced model is chosen. However, the above procedure has the drawback that there is no general way of incorporating the degree of evidence for or against the null hypothesis in the response estimator. Empirical Bayes estimators of $\underline{X}\underline{\beta}$ in the response function and $(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m)$ in the multivariate models are developed in this study. The structure of the corresponding prior distribution of $\underline{\beta}$ or $(\underline{\beta}_1, \cdots, \underline{\beta}_m)$ is motivated from the associated null hypothesis. The prior distribution in the multivariate regression model II is adopted from Reinsel (1985) and a new reduced model is proposed in Chapter 4. All the above priors are still the Zellner's (1986) g-priors by setting the prior variance-covariance matrix $\underline{\Sigma}$ $(\underline{\Sigma}_k,\ k = 1, \cdots, m)$ as $\tau^2(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}$ $(\tau^2(\underline{X}_k^T\underline{V}_k^{-1}\underline{X}_k)^{-1},\ k = 1, \cdots, m)$. The hierarchical Bayes estimators of $\underline{X}\underline{\beta}$ are only introduced in Chapter 2.

These empirical Bayes and hierarchical Bayes estimators are weighted averages of LSE's of $\underline{X}\underline{\beta}$ $(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m)$, based on the full model and the reduced model, or of the LSE based on the full model and a weighted average of LSE's for the reduced model (see multivariate regression model II). These weights are data dependent, and are such that the proposed estimators lean more towards the LSE of the full model if the F ratio for the corresponding hypothesis test is large, and to

the LSE of the reduced model, or the weighted average of LSE's of the reduced model, when the opposite is true.

Serveral conclusions relating to the EB estimators are summarized in the followings.

(1) Under a general quradratic loss, these empirical and hierarchical Bayes estimators are evaluated by their frequentist and Bayes risks. It has been shown that under certain conditions empirical Bayes estimators dominate the minimax estimator of $\underline{X}\underline{\beta}$ $(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m)$, the LSE of the full model, when comparisons are based on the frequentist criterion (see Corollary 2.2.1, Corollary 2.2.2, Corollary 3.2.1, and Corollary 4.2.1). The EB estimators then provides useful minimax estimators of $\underline{X}\underline{\beta}$ $(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m)$.

(2) If the quadratic loss is specified by $\underline{Q} = \underline{V}^{-1}$ $(\underline{Q}_k = \underline{V}_k^{-1})$, empirical Bayes estimators turn out to be fairly robust when compared using the frequentist risks to the LSE of the reduced model, and the weighted average of the LSE's of the reduced model.

(3) In Chapter 2, empirical Bayes estimators are also shown dominate the estimator of $\underline{X}\underline{\beta}$, which was proposed by Sclove (1968), in both the frequentist and Bayes senses (see Theorem 2.2.2 and Theorem 2.2.7).

(4) By using the argument in Theorem 6.2 of Lehmann (1983), it has been proved that the frequentist risk of the positive part of empirical Bayes estimators is less than that of empirical Bayes estimators (see Theorem 2.2.3, Theorem 3.2.2, and Theorem 4.2.2).

(5) The preliminary test estimators provide other compromises between the LSE of the full model and the LSE of the reduced model or the weighted average of the LSE's in the reduced model. For every preliminary test estimators of $\underline{X}\underline{\beta}$ $(\underline{X}_1\underline{\beta}_1, \cdots, \underline{X}_m\underline{\beta}_m)$, the corresponding modified empirical Bayes estimators

dominate the preliminary test estimators (see Theorem 2.2.4, Theorem 3.2.3, and Theorem 4.2.3).

(6) When the evaluation is based on the Bayes criterion, under a general quadratic loss, empircal Bayes estimators are superior to the LSE of the full model (see Theorem 2.2.5, Theorem 3.2.4, and Theorem 4.2.4).

(7) If the sampling variability is smaller than the prior variability, empirical Bayes estimators dominate the LSE of the reduced model or the weighted average of the LSE's for the reduced model. However, if the prior variability is relatively smaller than the sample variability, then the LSE of the reduced model or the weighted average of the LSE's for the reduced model can become a more approrate estimator (see Theorem 2.2.6, Theorem 3.2.5, and Theorem 4.2.5).

5.2 Future Studies

There are several open problems in this area that are worth further study. First, instead of using Zellner's (1986) g-prior, one can use other priors. An interesting question would be to study the robustness of the estimators developed under the choice of different priors. Finally, hierarchical Bayes estimators should also be developed in the multivariate regression model, and their robustness properties should be studied.

BIBLIOGRAPHY

Arnold, S. F., 1981. "The Theory of Linear Models and Multivariate Analysis," John Wiley & Sons Inc., New York.

Bancroft, T. A., 1944. "On Biases in Estimation due to the Use of Preliminary Tests of Significance," Ann. Math. Statist., 15, 190-204.

Baranchik, A. J., 1970. "A Family of Minimax Estimtors of the Mean of a Multivariate Normal Distribution," Ann. Statist., 41, 642-645.

Baranchik, A. J., 1973. "Inadmissibility of Maximum Likelihood Estimators in some Multiple Regression Problems with Three or More Indep. Variables," Ann. Statist., 1, 312-321.7-1136.

Berger, J., 1976. "Admissible Minimax Estimation of a Multivariate Normal Mean with Arbitrary Quadratic Loss," Ann. Statist., 4, 223-226.

Berger, J., 1982. "Selecting a Minimax Estimator of a Multivariate Normal Mean," Ann. Statist., 10, 81-92.

Berger, J., and L. R. Haff, 1983. "A Class of Minimax Estimators of a Normal Mean Vector for Arbitrary Quadratic Loss and Unknown Covariance Matrix," Statistics & Decisions, 1, 105-130.

Bock, M. E., 1975. "Minimax Estimations of the Mean of a Multivariate Normal Distribution," Ann. Statist., 3, 209-218.

Efron, B., and C. Morris, 1975. "Data Analysis using Stein's Estimator and its Generalizations," Journal of the American Statistical Association, 70, 311-319.

Efron, B., and C. Morris, 1976. "Families of Minimax Estimators of the Mean of a Multivariate Normal Distribution," Ann. Statist., 4, 11-21.

Ghosh, M., A. K. MD. E. Saleh, and P. K. Sen, 1987. "Empirical Bayes Subset Estimation in Regression Model," Dept. of Stat., U. of Florida, Gainesville, FL., Technical report no. 281.

James, W., and C. Stein, 1961, "Estimation with Quadratic Loss," Proc. Fourth Berkeley Symp. Math. Statist. Prob., 1, 361-379, U. of California Press, Berkeley, CA.

Jeffreys, H., 1967, "Theory of Probability," Oxford University Press, London.

Judge, G. G., R. C. Hill, and M. E. Bock, 1988. "Estimation of the Multivariate Normal Mean under Quadratic Loss," preprint.

Khuri, A. I., and J. A. Cornell, 1987. "Response Surfaces Designs and Analysis," Marcel Dekker, Inc., New York.

Lehmann, E. L., 1983. "Theory of Point Estimation," John Wiley & Sons. Inc..

Morris, C., 1977. "Interval Estimation for Empirical Bayes Generalizations of Stein's estimator," Proceedings of the Twenty-Second Conference on the Design of Experiments in Army Research Development and Testing, ARO Report 77-2.

Muth, J. F., 1961. "Rational Expectations and the Theory of Price Movements," Econometrica, 29, 315-335.

Myers, R. H., 1976. "Response Surface Methodology," Dept. of Stat., Virginia Tech., Blacksburg, VA., Distributed by Edwards Brothers, Inc., Ann-Arbor, Michigan.

Rao, C. R., 1973. "Linear Statistical Inference and its Application," 2nd Edition, Wiley, New York.

Reinsel, C. G., 1985. "Mean Squared Error Properties of Empirical Bayes Estimators in a Multivariate Random Effects General Linear Model," Journal of the American Statistical Association, 80, 642-650.

Sclove, S. L., 1968. "Improved Estimators for Coefficients in Linear Regression," Journal of the American Statistical Association, 63, 596-606.

Stein, C., 1956. "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution," Proc. Third. Berkeley Symp. Math. Statist. Prob., 1, 197-206, U. of California Press.

Stein, C., 1960. "Multiple Regression," Contributions to probatility and statistics, essays in honor of Harold Hotelling, edited by Olkin, I., Stanford U. Press, Stanford, CA.

Strawderman, W. E., 1971. "Proper Bayes Minimax Estimators of the Multivariate Normal Mean," Annals. of Mathematical Statistics, 42, 385-388.

Zellner, A., 1986. "On Assessing Prior Distribution and Bayesian Regression Analysis with G-prior Distribution," Bayesian inference and Decision techniques, edited by Goel, P. and Zellner, A., Elsevier Science Publishers B.V..

BIOGRAPHICAL SKETCH

Li-Chu Lee was born on March 1, 1957, in Taiwan, the Repbulic of China. She received her Bachelor of Business Science degree in communication and transportation management science from the National Cheng-Kung University, Taiwan, the Republic of China, in 1979. She then worked as a business officer in the governormental telephone company for twenty months. She was married to Mr. Li-Hwa Lin in July 1980. She came to the United States and joined her husband who was a graduate student at Oregon State University in April 1981. She entered the same University as a graduate student in 1982 and graduated with a master's degree in operations research in 1983. In August, 1984, she was accepted into the graduate school at the University of Florida in the Department of Statistics. She has worked as a teaching assistant as well as a research assistant while earning the Doctor of Philosophy degree.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Malay Ghosh, Chairman
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Michael A. DeLorenzo
Associate Professor of Dairy Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Ramon C. Littell
Professor of Statistics

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Science and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

May, 1989

Dean, Graduate School